

source 1. .327
 /organism="Pyrococcus furiosus"
 /strain="DSM 3638"
 /db_xref="taxon:2261"
 /clone="0028"
 /lab_host="E. coli"
 /note="Vector: 1-ZAP II; Site.1: Eco RI; Site.2: Eco RI; Genomic DNA was purified from P. furiosus cells and partially digested with Dra I, Eco RV, Hinc II, size selected, ligated to Eco RI linkers then cloned into the Eco RI site of 1-ZAP II, plasmid excision vector. Excision was performed in batch and individual clones retrieved by plating."

BASE COUNT 87 a 82 c 57 g 101 t

ORIGIN

Query Match 7.4%; Score 138.2; DB 9; Length 327;
 Best Local Similarity 68.3%; Pred. No. 1.7e-23;
 Matches 207; Conservative 0; Mismatches 93; Indels 3; Gaps 1;

QY 1045 ACCCTGAGCCGAGGACTTCGCTTCAAGACAAAGACCTGAGATAGCGCTGTAC 1104
 DB 303 AACTTGGACACAGTTCCTCGTTATTAAGCTAGACTGAGATGATGTTGAT 244
 QY 1105 CTCAGGCCGAGGATTAAAGAG--GAGAAGGCCCGGTGATGCTGCTCCACGCCGG 1161
 DB 243 ATGAAGCCCGAGCTCAGAGAGGAGAAAAAGCCCGGTATGATTTGTCATGGTGG 184
 QY 1162 CCGAAGGCGATGACGAGACCCGCTTGTCTACGAGATGACAGCTGATGCGCAAGGCG 1221
 DB 183 CCCAAGGCGAATGTACGTTATTAATCAAGTATGAAATCAATTAATGCGGCCAAGGG 124
 QY 1222 TACTACTGCTGCTGTGTAACCGCGCGGACGAGCGGTATAGGAGACTTCGGCTC 1281
 DB 123 TACTATATAGTTTACGTTATCCAGAGGAGCAATGATTAAGTAAAGACTTTCACAT 64
 QY 1282 CGCGTCTGTGAGAGGACTGCTTGTGAGACTTGTGAGCAATATAGCGGATCGAGGAG 1341
 DB 63 AGAGTATTAACAGAGACGAGACTGAGAGGACTTCCAGATATATTAAAGCGTATGAGGA 4
 QY 1342 TTC 1344
 DB 3 TTC 1

RESULT 2 803 bp mRNA linear EST 03-JAN-2002
 BM321042
 LOCUS
 DEFINITION rocketfeller.0.613 Mastigamoeba balamuthi lambda zap II library
 Mastigamoeba balamuthi cDNA similar to acidic ribosomal protein
 large P0, mRNA sequence.
 ACCESSION BM321042
 VERSION BM321042.1 GI:18055448
 KEYWORDS EST
 SOURCE Mastigamoeba balamuthi.
 ORGANISM Mastigamoeba balamuthi.
 Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
 REFERENCE 1 (bases 1 to 803)
 AUTHORS Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and
 Philippe, H.
 TITLE The analysis of 100 genes supports the grouping of three highly
 divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
 MEDLINE 21819461
 COMMENT Contact: Muller Miklos
 Laboratory of Biochemical Parasitology
 The Rockefeller University
 1230 York Avenue, New York, NY 10021, USA
 Email: mmuller@rockefeller.edu
 Insert length: 803 Std Error: 0.00
 POLYX=yes.

FEATURES
 source Location/Qualifiers
 1. 803
 /organism="Mastigamoeba balamuthi"
 /strain="ATCC 30984"
 /db_xref="taxon:108607"
 /clone_lib="Mastigamoeba balamuthi lambda zap II library"
 /note="syn: Phreatamoeba balamuthi"

BASE COUNT 155 a 282 c 236 g 127 t 3 others

Query Match 2.8%; Score 53.2; DB 13; Length 803;
 Best Local Similarity 50.0%; Pred. No. 0.028;
 Matches 133; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 1103 ACCCTGAGCCGAGGATTAAAGAGAGAGAGGCCCGGTGATGCTTGTCCACGGGGCG 1162
 DB 98 ACATGCCCTCGAGATCAACACAGGCGCAGGTGATCTCCACGAGTCCAGTGTCTCA 157
 QY 1163 CGAAGGCGATGACGAGACACCGCTTGTCTACGAGATGAGTGTATGCGGCAAGGCT 1222
 DB 158 AGAAGGCGACAAAGTCCGCGCTCCGAGGCGCCCTGCTGCGCAAGTCAACATCCGCC 217
 QY 1223 ACTACTGCTGCTTGTGAAACCGCGCGGAGAGAGCGCTATAGCGAAGACTTCGGCTCC 1282
 DB 218 CGTTCAATGACGCTTATCCCTCCACGCTGTCTACACAGGCTTGTCTACGACACG 277
 QY 1283 GCGTCTGTGAGAGAGACTGCTTGTGAGAGCTTGTGAGCAATTAAGCGCATCGAGAGT 1342
 DB 278 GCGTCTGTGACATGACGAGCGGCGGAGCTGTACAGAGTCTTGGCAAGGCGCTCGGCCA 337
 QY 1343 TCTTCAAGCTCGAAGCGGCGCGAC 1368
 DB 338 TCGCCGCTCGGCTCGGATCGGCG 363

RESULT 3 664 bp mRNA linear EST 26-FEB-2002
 BM645896
 LOCUS
 DEFINITION 17000687317518 A.Gam.ad.cdna1 Anopheles gambiae cDNA clone
 19600449655917 5', mRNA sequence.
 ACCESSION BM645896
 VERSION BM645896.1 GI:18945407
 KEYWORDS EST
 SOURCE African malaria mosquito.
 ORGANISM Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anopheles.

REFERENCE 1 (bases 1 to 664)
 AUTHORS Holt, R.A., Loh, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Charlab
 R., Collins, F.H., Venter, J.C. and Hoffman, S.L.
 TITLE Celera Anopheles gambiae EST project
 JOURNAL Unpublished (2002)
 COMMENT Contact: Holt R.A.
 Celera Genomics
 45 W. Gude Dr., Rockville, MD 20850, USA
 Tel: 2404533151
 Fax: 2404534580
 Email: HoltR@celera.com
 Plate: N01004A99 row: M column: 03
 Seq primer: M13 Reverse.

FEATURES
 source Location/Qualifiers
 1. 664
 /organism="Anopheles gambiae"
 /strain="RSP-ST (Reduced susc. to Permethrin - std.
 chromosome)"
 /db_xref="taxon:7165"
 /clone_lib="A.Gam.ad.cdna1"
 /dev_stage="Adult"
 /lab_host="DHI0b"
 /note="Vector: pSport1; Site.1: SalI; Site.2: NotI; Whole
 adult mosquitoes (mixed sex) frozen on liquid nitrogen."

adult mosquitoes (mixed sex) frozen on liquid nitrogen.


```

FEATURES
SOURCE
Location/Qualifiers
1. .453
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:480882"
/clone_1fb="Soares mouse p3NMF19.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/notes="Vector: pT733 (pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dn) primer [5',
TGTTACCAATCTGAAAGTGGCGGCGCGCATTTTCTTTTCTTTTCTTTT 3',
double-stranded cDNA was size selected, ligated to Eco RI
adapters (pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT733 vector
(pharmacia). Library went through one round of
normalization to a cot = 5. Library constructed by Bento
Soares and M.Fatima.Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."
BASE COUNT
66 a 135 c 149 g 102 t 1 others

```

Query Match	2.6%	Score 49.2;	DB 9;	Length 453;
Best Local Similarity	46.6%	Pred. No. 0.21;		
Matches 156; Conservative	0;	Mismatches 179;	Indels 0;	Gaps 0;

QY	1504	GACATPAGGGCTCTGTAGACGTGAGAGTCATCGGGCCAAATTCGTTAGACAAAGACAC	1565
Db	441	GAGAAAGAGCTGNNAGAGGACACCCCAAGGGCTTCGGCCGGAAAGCGCGCTACAGCAAGAAA	382
QY	1564	TTTCAGGAAGCTCAGCCCGCTGTTTTCAGCTCAGAACTGAAAGGGCGCGATCTCTTAATC	1623
Db	381	TCGAGCGCAAGATGGCAAGCCGACCCCGGGTAAAGCGCGGCAAGAAAGGGACCGCGAGCGG	322
QY	1624	CACTCGCTTGAGGACTACCGCTGTCCGCTCAGCAGACCTTATGTTTACAAAGCTGTC	1683
Db	321	CAGCTTTTCGAGAGCTCAGAGGCAAGCGCATCTCGCCCAACGTGCGGAGAGGCCAAGGCC	262
QY	1684	AAGGACATGGGCAAGGAAGCCTACATACCATATTCAAGCGCGCGGCCACAGGCCACAGC	1743
Db	261	ACCAGTCGCTCAACAGAGGCGCTTGCGCGCGGTGCGCAAGATCATCCCAAGCGTCCCTCT	202
QY	1744	GTCGCGGGAAGCCCGAGGACACAGGCGGAGCGCTACAGAGCTTCCTATAGAGTTCGAG	1803
Db	201	GACAAGCTACCAAGATCCAGACGCTCAAGCTGGGCGCGCAGGTACATGAGCTTCTCTAC	142
QY	1804	CGCAAGCTCAGAAGATACGAGGAGGCGTTTAGGT	1838
Db	141	CAGGTTCTCCAGAGCGAGATGGACATTAAGAT	107

RESULT 6	
BG228520	
LOCUS	
DEFINITION	BG228520 478 bp mRNA linear EST 08-FEB-2001 U44966.x1 Soares_NMAX_maxillary-process Mus musculus cDNA clone IMAGE:3513394 3' similar to TF:p97831 p97831 Dermo-1 PROTEIN., RNA sequence.
ACCESSION	BG228520
VERSION	BG228520.1 GI:12716024
KEYWORDS	EST. house mouse. Mus musculus
SOURCE	Eumariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ORGANISM	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)
JOURNAL	Contact: Robert Strausberg, Ph.D. Email: cgapsb-re@mail.nih.gov
COMMENT	This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

```

MG1:1393970
Trace considered overall poor quality
Possible reversed clone: similarity on Wrong strand
Possible reversed clone: polyT not found
High quality sequence stop: 1.
location/qualifiers
1..478
source

```

BASE COUNT	99 a	155 c	143 g	81 t
ORIGIN				

[illegible]

RESULT	7
AY108285	
LOCUS	AY108285
DEFINITION	AY108285 1341 bp mRNA
ACCESSION	Zea mays PCO126006 mRNA sequence.
VERSION	AY108285
KEYWORDS	AY108285.1 GI:21211363
SOURCE	HTC.
ORGANISM	Zea mays. Zea mays. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytas; Magnoliopsida; Liliopsida; Poales; Poaceae; PACC Clade; Panicoideae; Andropogoneae; Zea. 1 (bases 1 to 1341) Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V. Mize Mapping Project/Duront Consensus Sequences for Design of Overgo Probes Unpublished (2002) 2 (bases 1 to 1341)
TITLE	JOURNAL REFERENCE

AUTHORS Coe, E. C.
 TITLE Direct Submission
 JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
 FEATURES Location/Qualifiers
 source 1. 1341
 /organism="Zea mays"
 /db_xref="MaizeDB:637344"
 /db_xref="taxon:4577"
 /clone="PC0126006"
 /clone_11d="Maize Mapping Project/Dupont Consensus Library"
 /note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed Dupont contigs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACS in conjunction with the Maize Mapping Project"

BASE COUNT 232 a 410 c 443 g 250 t 6 others
 ORIGIN

Query Match 2.6%; Score 48; DB 11; Length 1341;
 Best Local Similarity 44.4%; Pred. No. 0.7;
 Matches 192; Conservative 0; Mismatches 240; Indels 0; Gaps 0;

QY 948 CACATGACCGCCACGAGATAGGCGGCTTCTCTACGACGGCGAGCTGAACAGGT 1007
 DB 341 CAAGAGCACCCTATGCCCTCTCGAGCGCTTCTACGACCGCTGCGGAAAGTGAT 400
 QY 1008 CACCATATACAGCGGCGCATATTCAGAGACTCAGACCTTGAGCGCGAGCTTCCG 1067
 DB 401 GATCGAGCGCAAGAGACATCGGCGCCCTGAACCTCAAGTCCCTCAAGATCGGGCT 460
 QY 1068 CTTCAGAGAGAAAGCTCGAGATAGACGGCTGTGTACCTAGCGCGAGGTAAAGAGA 1127
 DB 461 GGTGCAAGAGAGACCCGTGTCTTCCGACACCATCTCGAGACATGCGCCATAGCGGAG 520
 QY 1128 GAAGGCGCGGTATGTCTTCTCCACGGCGCGGCGGAGGCGATGAGACCGCTT 1187
 DB 521 GACGCGCGCAGCAGAGAGAGGTCTGTGAGGCGGCGCAAGTGGCCAGCTGCACGCGCTT 580
 QY 1188 CGTCTACGAGATGACGCTGATGCGGAGCAAGGGCTCTACTGCTGCTGTAAGCCGGG 1247
 DB 581 CGTACAGCGGCTCCCGAGCGGTACAGAGACCGCGGCGGAGCGGGGTGACGCTGTC 640
 QY 1248 CGGACAGCGGCTTATAGGAGAACTTCGCGCTCGCTCTGAGAGAGACTGGCTTGA 1307
 DB 641 CGCGGCGCAGAGACGACGATCGCCATCGCGCGCGGCTCAAGAGACCGCGCGTGT 700
 QY 1308 GGACTTGAAGACATATAGAGCGCATGAGAGGTCTTCAAGCTGGAACCGAGCGCA 1367
 DB 701 GCTGCTGAGAGAGCGCCAGCGCTGAGCGCGAGCTCGAGTGTCTGAGAGAGC 760
 QY 1368 CAGGAGCGCGT 1379
 DB 761 GCTGAGCGCAT 772

RESULT 8
 BE906698
 LOCUS BE906698 870 bp mRNA linear EST 20-OCT-2000
 DEFINITION 601500641F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3902482 5',
 mRNA sequence.
 ACCESSION BE906698
 VERSION BE906698.1 GI:10400041
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 870)
 AUTHORS NIH-MGC <http://mgi.mcl.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: L1M9705 row: a column: 11
 High quality sequence stop: 640.
 Location/Qualifiers
 source 1. 870
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3902482"
 /clone_11d="NIH_MGC_70"
 /tissue_type="epithelioid carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="organ: pancreas; Vector: pCMV-SPORT6; Site: 1; Note: Site 2: Salt. Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.1 kb. Library constructed by Life Technologies."

BASE COUNT 116 a 335 c 228 g 191 t
 ORIGIN

Query Match 2.6%; Score 47.8; DB 12; Length 870;
 Best Local Similarity 46.1%; Pred. No. 0.63;
 Matches 160; Conservative 0; Mismatches 187; Indels 0; Gaps 0;

QY 1062 CTTCGCTTCAGAGCAAGACCTCGATAGAGCGCTGTACTCAGCGGAGGTAA 1121
 DB 321 CATGCCCTTAATGGCCCTACTGAGAGCTTGTGCGACACTTATCAGCTGTGCT 380
 QY 1122 AGAGAGAAAGGCCCGGTATGATCTTGTCCAGCGCGGCGGAGGAGATGACGACA 1181
 DB 381 CGGCCACCTGTGTGCTCTCATGTCAAGCGCCGCTGCTGTATGTACTGCTCTA 440
 QY 1182 CGGCTGCTTACAGATGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1241
 DB 441 TCTCTTCTTCCGATGACAGCTGAGAGATTTCAAGGAGGAGGAGGAGGAGGAGGAG 500
 QY 1242 CCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1301
 DB 501 CTACTGTGAGTGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 560
 QY 1302 CTGAGAGGACTTGAAGACATATAGAGCGCATCGAGAGTCTTCAAGCTGGAACCGCA 1361
 DB 561 CCGGGGGCTGCTCCGCGCTCCATCGGCTACTTCTTCTTCCCTCCCATTTCTGTGT 620
 QY 1362 GGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1408
 DB 621 GCGTGGCTTGGCCCTGTGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 667

RESULT 9
 AL528820
 LOCUS AL528820 967 bp mRNA linear EST 13-FEB-2001
 DEFINITION AL528820 L1.NFL001.NBC4 Homo sapiens cDNA clone CS00D001YF08 5'
 prime, mRNA sequence.
 ACCESSION AL528820
 VERSION AL528820.1 GI:12792313
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 967)
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope

FEATURES

genoscope - Centre National de Sequençage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Location/Qualifiers

BASE COUNT	146 a	326 c	268 g	218 t	9 others
ORIGIN					

Query Match	2.68;	Score 47.8;	DB 9;	Length 967;
Best Local Similarity	49.08;	Pred. No. 0.67;		
Matches 121; Conservative	2;	Mismatches 124;	Indels 0;	Gaps 0;

QY	1062	CTTCCGGTTCAAGACAAAGACCTCGAGATAGACGGCTGGTACCTCAGGCCGGAGGTTAA	1122
Db	118	CATCCCTTGAAATTGGCCCTACCTGMAAGTCTCTGGCCAGACCTTCAATCACCGTGCCTKT	177
QY	1122	AGAGAGAGAGGCCCCGGGTGATAGTCTTCCTTCGCCAGCGCGGGCCGGAAGGGCATGTACGACA	1181
Db	178	YGGCACACTGGTCTGCTCTCAATGTACAGGGTCCCGTCCGTCTATGTATCACTGCCTCTA	237
QY	1182	CCGCTTCGTACACAGATGCAGCTGATGGCAGAGAGGGCTATCACTGCTGTGGTGA	1241
Db	238	TCTTTTCTTTCGGCACATGGCACAGCTGAGGAATTTCAAGGGCACTTACTCTTACCTTGTGCC	297
QY	1242	CCCGCGGGGCAGCAGCGGCTATPACGAGAACTTGCGCGCTCCGCGTCTCGAGAGAGACTGG	1301
Db	298	CTACTCGTGTGGCTTATGTGTGTGTGACACTCGGTGTGTATCCGTGGAGATYCACCCGG	357
QY	1302	CTTGGAG	1308
Db	358	CTTGGGG	364

RESULT	10
BMS548451	
LOCUS	
DEFINITION	BMS548451 1227 bp mRNA linear EST 20-FEB-2007
ACCESSION	AGNCOURT_6573227 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5731944
VERSION	5 , mRNA sequence.
KEYWORDS	BMS548451
SOURCE	BMS548451.1 GI:18763010 EST . human. Organism
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 1227) NIH-MGC http://mgc.ncl.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC)
AUTHORS	Unpublished (1999)
JOURNAL	Contact: Robert Strausberg, Ph.D. Email: cgabbs@mail.nih.gov Tissue Procurement: Invitrogen
COMMENT	cDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.N.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clonal distribution: MGC clonal distribution information can be

```

found through the I.M.A.G.E. Consortium/LLNL at
http://image.llnl.gov
Plate: LLNL2733 row: e column: 01
High quality sequence stop: 487.
Location/Qualifiers
1..1227

```

BASE COUNT	239 a	476 c	275 g	237 t
ORIGIN				

Query Match	2.5%;	Score 47.4;	DB 13;	Length 1227;
Best Local Similarity	46.7%;	Pred. No. 0.94;		
Matches 150;	Conservative 0;	Mismatches 171;	Indels 0;	Gaps 0;

QY	1062	CTTCCGGCTTCAAGAGCAAAAGACCTCCGAGATGAGCGGGCTACCTGACAGCCGGAGGTTAA	1122
Db	174	CATGCCTCTGAAATTGGCCCTACCTGAAAGGTCTTTGGCCAGACCTTCATCACCGTCTCTGT	233
QY	1122	AGAGAGAAAGCCCCCGGTGATAGTCTTCCTCCACGCGCGCGCGAAGGCAATGTACGACA	1182
Db	234	CGCGACCTGGTGTGCTCTCAATATGTCAGGGTCCCGGTGCTGTGTGTACTACCTGCTCTA	293
QY	1182	CCGCTTCTCTACGAGATGACGTAGTGGCAGCAAGGGCTACTACTGCTCTTCTGTGAA	1241
Db	294	TCTCTTCTTCCGACATGGACAGCTGAGGAATTTTCAAGGGCACTACTCTCTACTCTGTCTCC	353
QY	1242	CCCGCGGGCGGACGACGGCTTATACGGAAGATTTGGCGCTTCGCGCTCCGGAGAGGACATGG	1301
Db	354	CTACCTGGTGTGCTTCAATGTGTGTGACCTTCCGTGTCTTATCTGCTGGAGTCCACCGG	413
QY	1302	CTTGGAGACCTTTGAGCATATATATGAACGGCATGAGAGATGTTCTTCAACCTGAAACCGCA	1361
Db	414	CTGTGGGGGTGCTCGCGCGCTCCATACGGGTACTTCTCTTCTGTGGCCTCCCATCTCT	473
QY	1362	GGCGGACAGGAGGCGCTTGG	1382
Db	474	GGCGGCGGCGCTGGCCCTGG	494

RESULT	11
BFF772766	
LOCUS	BFF772766 464 bp mRNA linear EST 12-JAN-2001
DEFINITION	IIS-TIT00027-131200-330-f12.IT0027 Homo sapiens cDNA, mRNA sequence.
ACCESSION	BFF772766
VERSION	BFF772766.1 GI:12120666
KEYWORDS	EST,
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 464)
AUTHORS	Dias Neto,E., Garcia Correa,R., Verjovsky-Almeida,S., Britones,M.R... Nagal,M.A., da SILVA,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalhal,A.F., Matsukuma,A., Balal,G.S., Simpson,D.H., Brinheim,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare .M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.T. and Simpson,A.J.
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

IMAGE:3971384 3' similar to TR:P97831 P97831 DERMO-1, PROTEIN. ;
mRNA sequence.
BF721977
VERSION BF721977.1 GI:12022979
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 548)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/resources.shtml
MGI:1471416
Possible reversed clone: similarity on wrong strand
Possible reversed clone: polyT not found
High quality sequence stop: 490.
FEATURES
source
1. 548
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3971384"
/clone_lib="Soares.NMBA.branchial.arch"
/tissue="type="branchial arches"
/dev_stage="embryo, 10.5 dpc"
/lab_host="DH10B (phage resistant)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: NotI. Site 2: EcoRI. 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGGAGCGGCGGCGGCTATTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT 105 a 191 c 179 g 73 t
ORIGIN
Query Match 2.5%; Score 47; DB 12; Length 548;
Best Local Similarity 46.3%; Pred. No. 0.79;
Matches 155; Conservative 0; Mismatches 180; Indels 0; Gaps 0;
QY 1504 GACATAGGCTCTGTCAGACGTCATCGGGCCAAATCCGTTAGAGACGAGAAC 1563
123 GAGGAGAGCTGAGAGGAGCGCCAGCCCTTCGGCGGAGCGGCGCTACAGAGANA 182
QY 1564 TTCAGAACTGACCCCGCTGTTCTACGCTCAGAACGTGAGGCGCGTACTCTAATC 1623
183 TCGAGGGAAGATGCGACGCCGACCCCGGTTAAGCGCGCAAGAGGCGCGGAGCGCG 242
QY 1624 CACTCGCTTGAAGACTACCGCTGTCGCTGACACAGAGCCTTATGTTTACACAGTGTTC 1683
243 CAGCTCTTTCGAGAGAGCTGACAGACGCGCATCTCGGCCACAGGTGGGCGGCGAGCGCG 302
QY 1684 AAGGACATGGGCGAAGAACCTACATAGGATATTCACAGCGGCGCGCCACGCGACAGC 1743
303 ACCCACTGCTCAACAGAGCCTTCGCGCGCTGCGCAAGATCATCCCAAGCTCCCTCT 362
QY 1744 GTCCGCGGAGCCCGGAGCGAGCGCGCTACAGAGCCTTCTATAGATTCTTTGAG 1803
363 GACAGCTCAGCAAGATCCAGACGCTCAAGCTGGCGCGGCGGCTACATAGACTTCTCTAC 422
DB

QY 1804 CGCAGCTCAGAGATGAGAGGAGGCTTGTAGCT 1838
DB 423 CAGCTTCTCAGAGCGAGATGAGCAATAGAT 457
RESULT 14
BE585051
LOCUS BE585051 614 bp mRNA linear EST 16-AUG-2000
DEFINITION 7-7A-20 Psoljaezo Phytophthora sojae cDNA, mRNA sequence.
VERSION BE585051.1 GI:9836000
KEYWORDS EST.
SOURCE Phytophthora sojae.
ORGANISM Phytophthora sojae.
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
Phytophthora.
REFERENCE 1 (bases 1 to 614)
Qutob,D., Hrabec,P.F., Sobral,B.W.S. and Gijzen,M.
Comparative analysis of expressed sequences in Phytophthora sojae
JOURNAL Plant Physiol. 123 (1), 243-254 (2000)
MEDLINE 20267956
COMMENT Contact: Gijzen M
Agriculture and Agri-Food Canada
1391 Sandford Street, London, Ontario, Canada N5V 4T3
Tel: 519 457 1470
Fax: 519 457 3997
Email: gijzenm@agr.ca.
FEATURES
source
1. 614
Location/Qualifiers
/organism="Phytophthora sojae"
/strain="race 2, strain P6497"
/db_xref="taxon:67593"
/clone_lib="Psoljaezo"
/dev_stage="zoospores"
/lab_host="E. coli strain XL0LR"
/note="Vector: pBR-CMV. Site 1: EcoRI. Site 2: XhoI. This
cDNA library was constructed from polyA+ enriched mRNA
from zoospores grown in liquid medium. Zoospores were
released into water and collected by centrifugation at
2,000g; zoospore-bearing sporangia were induced to
develop on 5 to 7 d old mycelium colonies grown on V8
agar by repeated flooding with water. Complementary DNA
was synthesized from mRNA using an XhoI-poly(dT)
linker-primer. EcoRI adapters were ligated to the
blunt-ended cDNA fragments and the products were digested
with XhoI for directional cloning into lambda ZAP Express
vector. This lambda library was amplified once using E.
coli host strain XL1 Blue MRF'. Inserts were then
subcloned by mass excision using ExAssist helper phage
for conversion into phagemid vector pBR-CMV in E. coli
host strain XL0LR. Sequenced using T3 primer: 5' ATT AAC
CCT CAC TAA AGC GA 3'."
BASE COUNT 124 a 198 c 203 g 89 t
ORIGIN
Query Match 2.5%; Score 47; DB 10; Length 614;
Best Local Similarity 47.2%; Pred. No. 0.84;
Matches 177; Conservative 0; Mismatches 195; Indels 3; Gaps 1;
QY 1474 ACCTACTGGCTCACAGCTACGCTTCTCGACATAGGCGCTGTGACAGCTGAGTC 1533
136 ACCGTCCTGAGGAGGAGCTTCTCTCGCCACAGCGGCTCCGACCGGCGCTGTG 195
QY 1534 ATCGGGCCAAATCCGTTAGAGAGAACTTCAGAGAGCTAGCGCGCTGTTCAGCT 1593
136 ATGATGCGCCAGCGCTCCGCCCAATGAAAGCGCTGTGCAAGACCCGCTGTTCAG 255
QY 1594 CAGAAGGTGAAG---GCGCGCATCTCTAATCACTCACTGTTGAGAGTACCGCTGTCCG 1650
256 ACGGCGCTCAGCGCGGTGCAATGACAGACTTCTCGCGGAGATCAACCAATGAGAG 315
QY 1651 CTCGACAGAGCCTTATGTTTACAGAGTGTCTCAAGACATAGGCGCAAGAACTTACATA 1710
11111 1 1 11111 1 1 111

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2003, 14:01:35 ; Search time 319 Seconds

(without alignments)
9251.230 Million cell updates/sec

Title: US-09-966-803-1

Perfect score: 1869

Sequence: 1 ATGACGGCATCGATGGA.....TCAGGGGAATGGACTGA 1869

Scoring table:

IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1105431 segs, 789497651 residues

Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/PCUS_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/PCUS_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1869	100.0	1869	10	US-09-966-803-1
2	140.8	7.5	1958	10	US-09-974-300-224
3	66.2	3.5	477	10	US-09-974-300-4866
4	55	2.9	783	10	US-09-815-242-8012
5	52.2	2.8	2817	9	US-10-156-761-5055
6	52.2	2.8	9025608	9	US-10-156-761-5055
7	50	2.7	1446	9	US-10-156-761-2793
8	50	2.7	9025608	9	US-10-156-761-2793
9	48.2	2.6	1503	9	US-10-156-761-6210
10	47	2.5	1785	10	US-09-864-761-18040
11	47	2.5	2428	10	US-09-822-849A-6
12	47	2.5	2673	9	US-10-060-425-1
13	46.6	2.5	1431	9	US-10-156-761-2673
14	46.6	2.5	18438	9	US-10-156-761-2886
15	46.6	2.5	125746	9	US-10-156-761-15102
16	46.6	2.5	3372	9	US-10-067-457-2
17	45.8	2.5	504	9	US-09-918-995-26097
18	45.6	2.4	1792	9	US-09-086-436-40
19	45	2.4	999	9	US-09-925-664-46

20	44.6	2.4	489	10	US-09-864-761-1278	Sequence 1278, Ap
21	44.4	2.4	594	9	US-10-123-155-10	Sequence 10, Appl
22	44.4	2.4	924	9	US-10-156-761-3907	Sequence 3907, Ap
23	43.8	2.3	1941	9	US-10-156-761-3418	Sequence 3418, Ap
24	43.4	2.3	681	9	US-10-156-761-3307	Sequence 3307, Ap
25	43.2	2.3	1155	9	US-10-156-761-4992	Sequence 4992, Ap
26	43	2.3	1556	9	US-10-081-872-121	Sequence 121, Appl
27	42.8	2.3	1176	9	US-10-156-761-3069	Sequence 3069, Ap
28	42.8	2.3	1299	9	US-09-977-719-180	Sequence 180, Appl
29	42.8	2.3	1809	9	US-10-156-761-4041	Sequence 4041, Ap
30	42.8	2.3	1899	9	US-10-156-761-7216	Sequence 7216, Ap
31	42.8	2.3	2051	9	US-09-975-719-114	Sequence 114, Appl
32	42.8	2.3	3012	9	US-10-156-761-7256	Sequence 7256, Ap
33	42.6	2.3	1272	9	US-10-156-761-2079	Sequence 2079, Ap
34	42.6	2.3	2149	10	US-10-171-581-103	Sequence 103, Appl
35	42.6	2.3	2149	10	US-09-880-107-3031	Sequence 3031, Ap
36	42.4	2.3	1014	9	US-10-156-761-5381	Sequence 5381, Ap
37	42.4	2.3	1512	10	US-09-880-107-3452	Sequence 3452, Ap
38	42.4	2.3	1632	9	US-10-156-761-6045	Sequence 6045, Ap
39	42	2.2	675	9	US-10-156-761-2995	Sequence 2995, Ap
40	42	2.2	1164	9	US-10-156-761-2298	Sequence 2298, Ap
41	42	2.2	1830	9	US-10-081-872-181	Sequence 181, Appl
42	41.8	2.2	1362	9	US-10-156-761-6703	Sequence 6703, Ap
43	41.6	2.2	1047	9	US-10-156-761-1815	Sequence 1815, Ap
44	41.4	2.2	1083	9	US-10-156-761-1132	Sequence 1132, Ap
45	41.2	2.2	1818	10	US-09-815-242-7989	Sequence 7989, Ap

ALIGNMENTS

RESULT 1
US-09-966-803-1
Sequence 1, Application US/09966803
Patent No. US20020137185A1

GENERAL INFORMATION:
APPLICANT: Murphy et al.
TITLE OF INVENTION: Amidases
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: CABELLA, BYRNE, BAIN, GILFILLAN,
CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/966, 803
FILING DATE: 27-Sep-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/609, 570
FILING DATE: 30-Jun-2000
APPLICATION NUMBER: 09/427, 372
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Charles J. Herion
REGISTRATION NUMBER: 28, 019
REFERENCE/DOCKET NUMBER: 331400-53
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1744
TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1869 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE

TOPOLOGY: LINEAR
MOLECULE TYPE: DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-966-803-1

Query Match 100.0%; Score 1869; DB 10; Length 1869;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1869; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGACCGGCATCGAATGGAACACGACACCTTTCTAAGTTCGCTACCTGGCGACCCG 60
DB 1 ATGACCGGCATCGAATGGAACACGACACCTTTCTAAGTTCGCTACCTGGCGACCCG 60
QY 61 AGGATACGGGAACTTAATGCGGTACACCTGACGAGGCGCAACATGAAGGACAAAG 120
DB 61 AGGATACGGGAACTTAATGCGGTACACCTGACGAGGCGCAACATGAAGGACAAAG 120
QY 121 TACGAGACACGAGTGTGTTGAAGACCTTGAAGGCGCTCAAGGCGCTTCATCGAAG 180
DB 121 TACGAGACACGAGTGTGTTGAAGACCTTGAAGGCGCTCAAGGCGCTTCATCGAAG 180
QY 181 GCCCTCATGCGGAGGATTTGCGCGAGGCGCAAGAACTCGCTTCACCTGCTTAAAGAG 240
DB 181 GCCCTCATGCGGAGGATTTGCGCGAGGCGCAAGAACTCGCTTCACCTGCTTAAAGAG 240
QY 241 GAGAAAGAGGAGACGAGATATGGGTGGCCGATATCCAGACCTGAGCGCCCAAGAAAGTC 300
DB 241 GAGAAAGAGGAGACGAGATATGGGTGGCCGATATCCAGACCTGAGCGCCCAAGAAAGTC 300
QY 301 CTCTCAACTAATAAAAGTCCGCTCGATCGATGAGAAAGCAATTCAGAGAGACTTTAGTT 360
DB 301 CTCTCAACTAATAAAAGTCCGCTCGATCGATGAGAAAGCAATTCAGAGAGACTTTAGTT 360
QY 361 GTGGGCTTCAAGAGAGGAGGAGAGTCTGCTCTTGAAGAGAGAGTCCGCGTGG 420
DB 361 GTGGGCTTCAAGAGAGGAGGAGAGTCTGCTCTTGAAGAGAGAGTCCGCGTGG 420
QY 421 TTGACACATATGAGATCTTGTGATGAGAGAGAGAGAGTCTGAGGTCTTGAAGTGAAG 480
DB 421 TTGACACATATGAGATCTTGTGATGAGAGAGAGAGAGTCTGAGGTCTTGAAGTGAAG 480
QY 481 GCCGAGAGATATATGAGAGAGTCTGAGAGAGAGAGTCTGAGGTCTTGAAGTGAAG 540
DB 481 GCCGAGAGATATATGAGAGAGTCTGAGAGAGAGAGTCTGAGGTCTTGAAGTGAAG 540
QY 541 GATCGATATGTTGTAAGCTCCGACCGAGGAGGAGCAAGCTTCCCTCTTAAAGTTT 600
DB 541 GATCGATATGTTGTAAGCTCCGACCGAGGAGGAGCAAGCTTCCCTCTTAAAGTTT 600
QY 601 TACGACATATGCTATGGAAGAGAGGAGGAGAGAGAGTCTTGAAGAGGTCTCTTC 660
DB 601 TACGACATATGCTATGGAAGAGAGGAGGAGAGAGAGTCTTGAAGAGGTCTCTTC 660
QY 661 GAGGCGGTGACTCCGAGGAGAGAGAGTCTTGAAGAGGAGAGAGAGAGAGAGAGAG 720
DB 661 GAGGCGGTGACTCCGAGGAGAGAGAGTCTTGAAGAGGAGAGAGAGAGAGAGAGAG 720
QY 721 ATCAGGAGAGAGAGAGAGTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
DB 721 ATCAGGAGAGAGAGAGAGTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
QY 781 CCGCTGACAGCTTGTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
DB 781 CCGCTGACAGCTTGTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
QY 841 GCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB 841 GCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 901 CACTGAGATTTACGGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
DB 901 CACTGAGATTTACGGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
```

```
QY 961 ACGAGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
DB 961 ACGAGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
QY 1021 GGGCGGATATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
DB 1021 GGGCGGATATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
QY 1081 GAGCTGAGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
DB 1081 GAGCTGAGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
QY 1141 ATAGTCTTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1141 ATAGTCTTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1201 CAGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
DB 1201 CAGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1261 TATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
DB 1261 TATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
QY 1321 ATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 1321 ATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 1381 GGAATTAACGGGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
DB 1381 GGAATTAACGGGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
QY 1441 CTCTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
DB 1441 CTCTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
QY 1501 TCGGACATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
DB 1501 TCGGACATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
QY 1561 AACTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
DB 1561 AACTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
QY 1621 ATCCACTGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
DB 1621 ATCCACTGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
QY 1681 CTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
DB 1681 CTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
QY 1741 AGGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
DB 1741 AGGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
QY 1801 GAGCGCAAGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
DB 1801 GAGCGCAAGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
QY 1861 GGGAGACTGA 1869
DB 1861 GGGAGACTGA 1869
```

RESULT 2
US-09-974-300-224
Sequence 224, Application US/09974300
Patent No. US20020146721A1
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
APPLICANT: Clausen, Id Groth
TITLE OF INVENTION: Methods for Monitoring Multiple Gene

RESULT 5
US-10-156-761-5055
: Sequence 5055, Application US/10156761
: Publication No. US20030119018A1
: GENERAL INFORMATION:
: APPLICANT: OMURA, SATOSHI
: APPLICANT: IKEDA, HARUO
: APPLICANT: ISHIKAWA, JUN
: APPLICANT: HORIKAWA, HIROSHI
: APPLICANT: SHIDA, TADAYOSHI
: APPLICANT: SAKAKI, YOSHIYUKI
: APPLICANT: HATTORI, MASAHIRU
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-262
: CURRENT APPLICATION NUMBER: US/10/156,761
: CURRENT FILING DATE: 2002-05-29
: PRIOR APPLICATION NUMBER: JP 2001-204089

```

RESULT 6
US-10-156-761-1
Sequence 1, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIOYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: misc_feature

```


[illegible]

```

: RESULT 10
: US-09-864-761-18040
: Sequence 18040, Application US/09864761
: Patent No. US20020048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
: FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
: FILE REFERENCE: Aecm1ca-x-1
: CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
: SEQ ID NO 18040
: LENGTH: 1785
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:

```


Query Match	2.58;	Score 47;	DB 10;	Length 2428;
Best Local Similarity	49.48;	Pred. No. 0.00052;		
Matches 122; Conservative	0;	Mismatches 125;	Indels 0;	Gaps 0;

Query Match	2.5%	Score 47	DB 94	Length 2673
Best Local Similarity	49.4%	Pred. No. 0.00054		
Matches 122	Conservative 0	Mismatches 125	Indels 0	Gaps 0
QY	1062	CTTCCGCTTCAGAGCAAGACCTCGAGATAGAGCGCTGTAACCTCAGACCGGAGGTTAA	11221	
Db	1419	CATGCCCTTGATATGGCCCTCAGAAAGGTCTTTGGCAGACCTTCAACACGTCCTGT	14768	
QY	1122	AGAGGAGAAGGCCCCGGTGATAGCTTCCTGTCACAGGCGGGCGAAGGGCATTGACGACA	11881	
Db	1479	CGGCACTAGTGTCTCCTCAATGTTCACAGCTCCCGTGCCTGTATGTCTTACCTGCTCTA	15338	
QY	1182	CCGCTTCGTCTACGAGATGCAGCTGATGGCGAGACAAGGGCTACTACTGCTCTTCGTGAA	12414	
Db	1539	TCCTTCCTTCGACATGGCACAGCTGAGAGAAATTTCAAGGGCACACTACTGCTTACCTGTCC	15998	
QY	1242	CCGCGCGGAGACACAGGCTTATACGAGAAGATTGCGCGTCCGCGCTGGAGAGAGACTGG	13018	
Db	1599	CTACCTGGTGTGCTTACATGTGTGTAGAGCTTCCGCTGTCACTGCTGGAGTCCACCGG	16588	
QY	1302	CTTGGAG	1308	
Db	1659	CTTGGG	1665	

RESULT 3
US-10-156-761-2673
; Sequence 2673, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO

QY	1306	GAGAGCTTTAGGACATTAAGACGGCATCGAGAGTTCCTTAAGCTCGAACCGAGGCC	1365
Db	42045	GCGGCGCTCCACGCTCGTACTGCACTCGCTCGCGGCGAGTTACCGACGCCCTCGCTGGCGC	42104
QY	1366	GACAGGAGAGCGCGCTGG	1382
Db	42105	GCAGTCCGCGCGCGGGG	42121

```
Search completed: July 15, 2003, 16:41:38
Job time : 334 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2003, 12:54:09 ; Search time 93 Seconds

(without alignments)
6163.214 Million cell updates/sec

Title: US-09-966-803-1

Sequence: 1 ATGACCGGATGCAATGGA.....TCAGGGGGAATGGCACTGA 1869.

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1869	100.0	1869	2 US-08-664-646A-1	Sequence 1, App1
2	1869	100.0	1869	2 US-09-066-285-1	Sequence 1, App1
3	1869	100.0	1869	3 US-09-261-006-1	Sequence 1, App1
4	1869	100.0	1869	3 US-08-951-088-1	Sequence 1, App1
5	1869	100.0	1869	4 US-09-609-566-1	Sequence 1, App1
6	1869	100.0	1869	4 US-09-609-570-1	Sequence 1, App1
7	120.4	6.4	1896	3 US-09-016-080-2	Sequence 2, App1
8	48.2	2.6	1598	1 US-08-211-682-24	Sequence 24, App1
9	46.4	2.5	5977	3 US-09-024-020B-1	Sequence 1, App1
10	46.4	2.5	5977	3 US-09-425-043-1	Sequence 1, App1
11	46.4	2.5	6007	4 US-09-024-020B-2	Sequence 2, App1
12	46.4	2.5	6007	4 US-09-425-043-2	Sequence 2, App1
13	46.4	2.5	6556	3 US-09-024-020B-7	Sequence 7, App1
14	46.4	2.5	6556	4 US-09-425-043-7	Sequence 8, App1
15	46.4	2.5	6586	4 US-09-024-020B-43	Sequence 43, App1
16	46.4	2.5	6586	4 US-09-425-043-43	Sequence 43, App1
17	46.4	2.5	6826	3 US-09-024-020B-8	Sequence 8, App1
18	46.4	2.5	6826	4 US-09-425-043-8	Sequence 8, App1
19	46.2	2.5	7218	1 US-08-232-463-14	Sequence 14, App1
20	42.8	2.4	999	2 US-08-761-277A-46	Sequence 46, App1
21	42.8	2.3	1299	4 US-09-199-637A-160	Sequence 160, App1
22	42.8	2.3	2051	4 US-09-199-637A-114	Sequence 114, App1
23	42.8	2.2	2241	2 US-08-838-219B-20	Sequence 20, App1
24	42.8	2.2	2241	3 US-09-233-336A-20	Sequence 20, App1
25	42.8	2.2	2241	3 US-09-233-752A-20	Sequence 20, App1
26	42.8	2.2	2241	4 US-09-402-036-20	Sequence 20, App1
27	42.8	2.2	2241	4 US-09-904-226-20	Sequence 20, App1

C	28	41.4	2.2	4403765	4 US-09-103-840A-2	Sequence 2, App1
C	30	41.4	2.2	4411529	4 US-09-103-840A-1	Sequence 1, App1
C	29	41.2	2.2	12588	2 US-07-951-715A-6	Sequence 6, App1
C	31	40.6	2.2	3624	1 US-08-459-448A-6	Sequence 6, App1
C	32	40.6	2.2	3624	2 US-08-459-595A-6	Sequence 6, App1
C	33	40.6	2.2	3624	3 US-08-459-504B-6	Sequence 6, App1
C	34	40.6	2.2	3624	3 US-08-459-444-6	Sequence 6, App1
C	35	40.6	2.2	3624	3 US-09-053-549-7	Sequence 7, App1
C	36	40.6	2.2	3624	3 US-09-547-422-6	Sequence 6, App1
C	37	40.6	2.2	1291	4 US-08-997-897-1	Sequence 1, App1
C	38	40.2	2.2	1291	4 US-09-156-836B-1	Sequence 1, App1
C	39	40.2	2.2	2852	4 US-09-056-556-203	Sequence 203, App1
C	40	40.2	2.2	2852	4 US-09-072-596-198	Sequence 198, App1
C	41	40.2	2.1	6344	4 US-08-843-417-1	Sequence 1, App1
C	42	40.0	2.1	6524	4 US-08-659-656A-1	Sequence 1, App1
C	43	40.0	2.1	6527	4 US-08-659-656A-7	Sequence 7, App1
C	44	40.0	2.1	7052	4 US-08-669-656A-5	Sequence 5, App1
C	45	40.0	2.1	7052	4 US-08-669-656A-5	Sequence 5, App1

ALIGNMENTS

RESULT 1
US-08-664-646A-1
Sequence 1, Application US/08664646A
Patent No. 5877001
GENERAL INFORMATION:
APPLICANT: Murphy et al.
TITLE OF INVENTION: Amides
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: CARELLA, BYRNE, BAIN, GIUFFILIAN,
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/664,646A
FILING DATE: June 17, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Charles J. Herron
REGISTRATION NUMBER: 28,019
REFERENCE/DOCKET NUMBER: 331400-53
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1744
TELEFAX: 201-994-1700
INFORMATION FOR SEQ. ID NO. 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1869 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: DNA
US-08-664-646A-1
Query Match 100.0%; Score 1869; DB 2; Length 1869;
Best local Similarity 100.0%; Pred. No. 0;
Matches 1869; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ATGACCGGATGCAATGGAACCAAGACCTTTTCTAAGTGGCTTGGGACCGG 60
|||||

Dp	1	ATGACCGGCATCGAATGGAAACACAGACACTTTTCTTAAGTTCCGCTACCTGGGACCCG	60
Oy	61	AGGATACGGGGAAACTTAATCGCGGTACACCTTGACGAAGGCCACATBAAGACAACAG	120
Dp	61	AGGATACGGGGAAACTTAATCGCGGTACACCTTGACGAAGGCCACATBAAGACAACAG	120
Oy	121	TACGAGACACGSGTTGTTGTGAAGACCTTGAACGGGCTCAAGGCGCTTATCGAGAAC	180
Dp	121	TACGAGACACGSGTTGTTGTGAAGACCTTGAACGGGCTCAAGGCGCTTATCGAGAAC	180
Oy	181	GCCTCAATGCCGAGAGATTTGCCACAGACGGCAGAAAAGCTCGCCTTCACCTGCTTAAACAG	240
Dp	181	GCCTCAATGCCGAGAGATTTGCCACAGACGGCAGAAAAGCTCGCCTTCACCTGCTTAAACAG	240
Oy	241	GAGAGAAGAGAGACCCGAGATATGAGGTGGCCGATATCCAGACCTCGAGGCGCCAAAGAAAGTC	300
Dp	241	GAGAGAAGAGAGACCCGAGATATGAGGTGGCCGATATCCAGACCTCGAGGCGCCAAAGAAAGTC	300
Oy	301	CTCTCAACTAAAAAGTCGCCGCTCGAGTCAGTCAGTGGAAACGACGATTTAAGGAGACTTTAGTT	360
Dp	301	CTCTCAACTAAAAAGTCGCCGCTCGAGTCAGTCAGTGGAAACGACGATTTAAGGAGACTTTAGTT	360
Oy	361	GTCCGCTTCAAGAGGAGGAGATGAGGACTTCGCTTGTGACGACAGCTGCTCCGCTCGG	420
Dp	361	GTCCGCTTCAAGAGGAGGAGATGAGGACTTCGCTTGTGACGACAGCTGCTCCGCTCGG	420
Oy	421	TTTCGACAATATGGGATTTCTTTGATGTGAGAGAGAAGACGACGTTCTGGGTTCTTGAACATGAG	480
Dp	421	TTTCGACAATATGGGATTTCTTTGATGTGAGAGAGAAGACGACGTTCTGGGTTCTTGAACATGAG	480
Oy	481	GCCGAGGAGATATATCGACACTTTTGAGAGAAGCCGAGGTTTGTGAGTGGCTTGGACAGGC	540
Dp	481	GCCGAGGAGATATATCGACACTTTTGAGAGAAGCCGAGGTTTGTGAGTGGCTTGGACAGGC	540
Oy	541	GATCGATAGTGTGTGAAGCTCCGCGACCGCAGAGGGGACCAAGCCGCTTGTCAAGTTC	600
Dp	541	GATCGATAGTGTGTGAAGCTCCGCGACCGCAGAGGGGACCAAGCCGCTTGTCAAGTTC	600
Oy	601	TACGACATATGTCTATGTGAGAGACGGGGAGAGAAAGCTCTTGACAGGSGTCTCTTC	660
Dp	601	TACGACATATGTCTATGTGAGAGACGGGGAGAGAAAGCTCTTGACAGGSGTCTCTTC	660
Oy	661	GAGCGGTTGACTCCGAGGGAAGGATATCTCGAGGGGCAAGAAAAAAGCGGTTTC	720
Dp	661	GAGCGGTTGACTCCGAGGGAAGGATATCTCGAGGGGCAAGAAAAAAGCGGTTTC	720
Oy	721	ATCAGACGACGACGACTGGCTGTACCTCTGGAGCGGCGAGCTTAAACCGATCTACAGAGGC	780
Dp	721	ATCAGACGACGACGACTGGCTGTACCTCTGGAGCGGCGAGCTTAAACCGATCTACAGAGGC	780
Oy	781	CCGCTCGACGCTGGGAAGCCAAAGCTCAAGGAAGAAAGGCTACTTCCCTACATCCAGAT	840
Dp	781	CCGCTCGACGCTGGGAAGCCAAAGCTCAAGGAAGAAAGGCTACTTCCCTACATCCAGAT	840
Oy	841	GCGGGCAAGGTAACCTCTGGCTCTGGAGACGGGAAGGCGCAGCGGTGTGTAAACGGCGAC	900
Dp	841	GCGGGCAAGGTAACCTCTGGCTCTGGAGACGGGAAGGCGCAGCGGTGTGTAAACGGCGAC	900
Oy	901	CACATGATTTACGGGCTTGACGTCACGAGATGGCAAAAGCATTTGCTCTCATATGACCGCC	960
Dp	901	CACATGATTTACGGGCTTGACGTCACGAGATGGCAAAAGCATTTGCTCTCATATGACCGCC	960
Oy	961	ACGAGAGATAGGCGAGCTCTACCTCTACGACGGCGAGCTGAAACAGGTACCCGAATACAC	1020
Dp	961	ACGAGAGATAGGCGAGCTCTACCTCTACGACGGCGAGCTGAAACAGGTACCCGAATACAC	1020
Oy	1021	GGGCGCATATTCAGGAAGCTCAAGACCTTCAGCGGAGGCACTTCGCTTCAAGGCAAA	1080
Dp	1021	GGGCGCATATTCAGGAAGCTCAAGACCTTCAGCGGAGGCACTTCGCTTCAAGGCAAA	1080
Oy	1081	GACCTCGAGATAGACGCGCTGGTACTCTCAGGCGCGAGGTTAAAGAGAGAAAGCCCGGTC	1140
Dp	1081	GACCTCGAGATAGACGCGCTGGTACTCTCAGGCGCGAGGTTAAAGAGAGAAAGCCCGGTC	1140
Oy	1140	GACCTCGAGATAGACGCGCTGGTACTCTCAGGCGCGAGGTTAAAGAGAGAAAGCCCGGTC	1200

```

OY 1141 ATATGCTTCGTCACACGGCGGGCCGAAGGGCATATGACGACACACGGCTTCGCTACGAGATG 1200
Db 1141 ATATGCTTCGTCACACGGCGGGCCGAAGGGCATATGACGACACACGGCTTCGCTACGAGATG 1200
OY 1201 CAGCTGATGCGGACCAAGGGGCTACTACGCTGCTTCGTAACCCGCGCCGACGAGCGGC 1260
Db 1201 CAGCTGATGCGGACCAAGGGGCTACTACTGCTGCTTCGTAACCCGCGCGGACGAGCGGC 1260
OY 1261 TATAGCGAAGACTTCGCGCTCCGCGTCCTGAGAGGACTGCGCTTGAGAGACTTGTAGAGAC 1320
Db 1261 TATAGCGAAGACTTCGCGCTCCGCGTCCTGAGAGGACTGCGCTTGAGAGACTTGTAGAGAC 1320
OY 1321 ATATATGAACGGGCAATCGAGGAGACTTCGTAACGCTGCAACCCGCGCGCCGACAGGAGGCGGT 1380
Db 1321 ATATATGAACGGGCAATCGAGGAGACTTCGTAACGCTGCAACCCGCGCGCCGACAGGAGGCGGT 1380
OY 1381 GGAATTAACGGGCAATAGGCTACGGGCGGCTCATGACCAACTGCGGCTTGACTACGACAGGCGAC 1440
Db 1381 GGAATTAACGGGCAATAGGCTACGGGCGGCTCATGACCAACTGCGGCTTGACTACGACAGGCGAC 1440
OY 1441 CTCTTCAAGCGAGCATTAAGCGAGACGGCATTAAGCTACTGCGCTCACAGCTACGCTTC 1500
Db 1441 CTCTTCAAGCGAGCATTAAGCGAGACGGCATTAAGCTACTGCGCTCACAGCTACGCTTC 1500
OY 1501 TCGGACATATAGGGCTCTGGTATACGAGCTGAGGTCAATCGGGCCCAATTCGCTTAGAGAACGAG 1560
Db 1501 TCGGACATATAGGGCTCTGGTATACGAGCTGAGGTCAATCGGGCCCAATTCGCTTAGAGAACGAG 1560
OY 1561 AACTTCAAGGAAGCTCAGGCCGCGCTGTTCACGCTCAGAACGTGAAGGGCGCGATACTCTTA 1620
Db 1561 AACTTCAAGGAAGCTCAGGCCGCGCTGTTCACGCTCAGAACGTGAAGGGCGCGATACTCTTA 1620
OY 1621 ATTCACATCGCTTAGAGGACTACCGCTGTCCGCTGCACACGAGGCTTAAGTCTTCAACAGTG 1680
Db 1621 ATTCACATCGCTTAGAGGACTACCGCTGTCCGCTGCACACGAGGCTTAAGTCTTCAACAGTG 1680
OY 1681 CTCAGAGGACATGGGCAAGGAAGGCTTACATACGATATTCAAGCGCGGCGCCACGCGCAC 1740
Db 1681 CTCAGAGGACATGGGCAAGGAAGGCTTACATACGATATTCAAGCGCGGCGCCACGCGCAC 1740
OY 1741 AGCGTCGCGGGAAGCCGAGGCGACAGGCGCGTACAGGCTTCCTCATAGAGTCTTTC 1800
Db 1741 AGCGTCGCGGGAAGCCGAGGCGCGTACAGGCGCGTTCCTCATAGAGTCTTTC 1800
OY 1801 GAGCGCAAGCTCAAGAAAGTACGAGAGGCGCTTGAAGTAGAAGATACCAAGGGGAAT 1860
Db 1801 GAGCGCAAGCTCAAGAAAGTACGAGAGGCGCTTGAAGTAGAAGATACCAAGGGGAAT 1860
OY 1861 GGGAACTGA 1869
Db 1861 GGGAACTGA 1869

RESULT 2
US-09-066-285-1
; Sequence 1, Application us/09066285
; Patent No. 5985646
; GENERAL INFORMATION:
; APPLICANT: Murphy et al.
; TITLE OF INVENTION: Amlidases
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELIA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
;

```



```

OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/066,285
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/664,646
FILING DATE: June 17, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Charles J. Hepton
REGISTRATION NUMBER: 28,019
REFERENCE/DOCKET NUMBER: 331400-53
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1869 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: DNA
US-09-066-285-1

```

```

Query Match      100.0%; Score 1869; DB 2; Length 1869;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1869; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ATGACGGGCAATGGAACACAGAGACCTTTCTAAGTTCGCTTACCTGGGCGACCCG 60
DB 1 ATGACGGGCAATGGAACACAGAGACCTTTCTAAGTTCGCTTACCTGGGCGACCCG 60
QY 61 AGGATACGGGGAACCTTAATCGGCTACACCTTACGAAAGGCCAACAATGAACAAG 120
DB 61 AGGATACGGGGAACCTTAATCGGCTACACCTTACGAAAGGCCAACAATGAACAAG 120
QY 121 TACGAGACACGGTGTGTGTTGAGACCTTGAAGAGGGCTCAAGGGGCTTCATCGAAG 180
DB 121 TACGAGACACGGTGTGTGTTGAGACCTTGAAGAGGGCTTCATCGAAG 180
QY 181 GCCTCAATGCCAGATTTGCGCAGACGGCAGAAAGCTGCTTCACTGCTTTAAACGAG 240
DB 181 GCCTCAATGCCAGATTTGCGCAGACGGCAGAAAGCTGCTTCACTGCTTTAAACGAG 240
QY 241 GAGAAAGAGAGCGAGATATGGGTGGGCGGATATCCAGACCTTGAGCCCAAGAAAGTC 300
DB 241 GAGAAAGAGAGCGAGATATGGGTGGGCGGATATCCAGACCTTGAGCCCAAGAAAGTC 300
QY 301 CTCTCAACTAAAAACCTCCGCTCGATGACAGTGAAGACGATTCAGAGAGACTCTAAGTT 360
DB 301 CTCTCAACTAAAAACCTCCGCTCGATGACAGTGAAGACGATTCAGAGAGACTCTAAGTT 360
QY 361 GTCGGGTTCAAGAGAGAGAGACGATGAGACCTTCTTTGACGACGACGCTCCGGTCTGG 420
DB 361 GTCGGGTTCAAGAGAGAGAGACGATGAGACCTTCTTTGACGACGACGCTCCGGTCTGG 420
QY 421 TTCGCAATATGGGATTTCTTGATGAGAGAGAGACGACCTTCTGGGTTTGAACACTGAG 480
DB 421 TTCGCAATATGGGATTTCTTGATGAGAGAGAGACGACCTTCTGGGTTTGAACACTGAG 480
QY 481 GCCGAGAGATATGAGAGAGTTCGAGAGACCGAGGTTTTCAGATGGGCTCTGGACGAGC 540
DB 481 GCCGAGAGATATGAGAGAGTTCGAGAGACCGAGGTTTTCAGATGGGCTCTGGACGAGC 540
QY 541 GATCGGATATGTTGTAAGCTCCCGCAGCGGAGGGGAGCAGACCTGCTGTTCAAGTTTC 600
DB 541 GATCGGATATGTTGTAAGCTCCCGCAGCGGAGGGGAGCAGACCTGCTGTTCAAGTTTC 600
QY 601 TACGACATATGCTTATGGAAGAGCGGGGAGAGAGAAAGCTTCTGAGAGGGTCTTCCTTC 660
DB 601 TACGACATATGCTTATGGAAGAGCGGGGAGAGAGAAAGCTTCTGAGAGGGTCTTCCTTC 660

```

```

QY 661 GAGCGGTTGATCTCCGAGCGAAGAGATATCTCTGAGGGGCAAGAAAAAAGCGGTTTC 720
DB 661 GAGCGGTTGATCTCCGAGCGAAGAGATATCTCTGAGGGGCAAGAAAAAAGCGGTTTC 720
QY 721 ATCAGAGAGACACATGCTGGCTGTACTCTGGGAGGGGAGCTTAAACCGATCTACGAGGC 780
DB 721 ATCAGAGAGACACATGCTGGCTGTACTCTGGGAGGGGAGCTTAAACCGATCTACGAGGC 780
QY 781 CCGCTGACGCTGTGGGAAGCCAAAGCTACAGGAAGAAAGGTTACTTCTCACTCCAGAT 840
DB 781 CCGCTGACGCTGTGGGAAGCCAAAGCTACAGGAAGAAAGGTTACTTCTCACTCCAGAT 840
QY 841 GCGGCGAGGGTAAACCTCTGCTGTGGAGCGGAGCGGAGCTGTTTACCGGCGAC 900
DB 841 GCGGCGAGGGTAAACCTCTGCTGTGGAGCGGAGCGGAGCTGTTTACCGGCGAC 900
QY 901 CACTGATATGAGGGCTGTGACGTACAGGATGCGAAGGCAATGCTCTCATCATGACCGCC 960
DB 901 CACTGATATGAGGGCTGTGACGTACAGGATGCGAAGGCAATGCTCTCATCATGACCGCC 960
QY 961 ACAGAGATAGCGAGCTCTACCTCTACGACGCGGAGCTGAACAGGTCAACGAAATCAAC 1020
DB 961 ACAGAGATAGCGAGCTCTACCTCTACGACGCGGAGCTGAACAGGTCAACGAAATCAAC 1020
QY 1021 GGGCCGATATTCAGGAAGCTCAAGACCTTCGACCGAGCGACTCTCCGTTCAAGACCAAA 1080
DB 1021 GGGCCGATATTCAGGAAGCTCAAGACCTTCGACCGAGCGACTCTCCGTTCAAGACCAAA 1080
QY 1081 GACCTGAGATATAGCGGCTGTACTCAGGCGGAGGTTAAAGAGAGAAAGGCCCGGGT 1140
DB 1081 GACCTGAGATATAGCGGCTGTACTCAGGCGGAGGTTAAAGAGAGAAAGGCCCGGGT 1140
QY 1141 ATAGTCTTCTCCACGCGCGGCGGAGGAGCATGTACGACACCGCTTCTCTACAGAGAT 1200
DB 1141 ATAGTCTTCTCCACGCGCGGCGGAGGAGCATGTACGACACCGCTTCTCTACAGAGAT 1200
QY 1201 CAGCGTATGCGGAGCAAGGGCTACTACTGCTGTTGTAACCCGCGGCGACCGACGCG 1260
DB 1201 CAGCGTATGCGGAGCAAGGGCTACTACTGCTGTTGTAACCCGCGGCGACCGACGCG 1260
QY 1261 TATAGGAGACTTCCGCTCCGCTCGCTGAGAGAGACGAGGCTTGGAGAGACTTGGAGAC 1320
DB 1261 TATAGGAGACTTCCGCTCCGCTCGCTGAGAGAGACGAGGCTTGGAGAGACTTGGAGAC 1320
QY 1321 ATATAGACGCGATCGAGAGATTTCAAGCTCGAACCAGCGAGCGCAGAGAGCGCGTT 1380
DB 1321 ATATAGACGCGATCGAGAGATTTCAAGCTCGAACCAGCGAGCGCAGAGAGCGCGTT 1380
QY 1381 GGAATTAAGGGGATTAAGCTACGCGGCTTCATGACCAACTGGGCTGTGACTAGAGCGAC 1440
DB 1381 GGAATTAAGGGGATTAAGCTACGCGGCTTCATGACCAACTGGGCTGTGACTAGAGCGAC 1440
QY 1441 CTCTTCAAGGAGATTAAGAGAGAGAGAGCATATAGCTATGCTACACGATACGCTTC 1500
DB 1441 CTCTTCAAGGAGATTAAGAGAGAGAGAGCATATAGCTATGCTACACGATACGCTTC 1500
QY 1501 TCGGACATATGGGCTGTGTAAGAGCTGAGAGCTATCGGGCCAAATCCGTTAGAGAAAG 1560
DB 1501 TCGGACATATGGGCTGTGTAAGAGCTGAGAGCTATCGGGCCAAATCCGTTAGAGAAAG 1560
QY 1561 AACCTCAGGAGCTCAGCGCGCTGTCTTCAAGCTCAGAAAGCTGAAGGCCCGATATCTCT 1620
DB 1561 AACCTCAGGAGCTCAGCGCGCTGTCTTCAAGCTCAGAAAGCTGAAGGCCCGATATCTCT 1620
QY 1621 ATCCACTGCTTGAAGACTACCGCTGTGCTCGACAGAGGCTTATGTTCTACAACTG 1680
DB 1621 ATCCACTGCTTGAAGACTACCGCTGTGCTCGACAGAGGCTTATGTTCTACAACTG 1680
QY 1681 CTCAAGGACATGCGCAAGAGACCTTACATATGAGATATTCAGAGCGCGCCACGCGCAC 1740
DB 1681 CTCAAGGACATGCGCAAGAGACCTTACATATGAGATATTCAGAGCGCGCCACGCGCAC 1740
QY 1741 AGCGTCCCGGAAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800
DB 1741 AGCGTCCCGGAAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800

```


1741 AGGTCGGGGAAGCCCGAGGCGACAGGCCGACAGGCTCTTCAAGAGTCTTC 1800
QY 1801 GAGCGCAAGCTCAAGAGTACGAGAGGCTTTAGAGTACAGAACTCAAGGGAAAT 1860
DB 1801 GAGCGCAAGCTCAAGAGTACGAGAGGCTTTAGAGTACAGAACTCAAGGGAAAT 1860
QY 1861 GGGAACTGA 1869
DB 1861 GGGAACTGA 1869

RESULT 3

US-09-261-006-1
Sequence 1, Application US/09261006

Patent No. 6004796
GENERAL INFORMATION:
APPLICANT: Murphy et al.
TITLE OF INVENTION: Amlases
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/261, 006
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/664, 646
FILING DATE: June 17, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Charles J. Herion
REGISTRATION NUMBER: 28,019
REFERENCE/DOCKET NUMBER: 331400-53
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ. ID NO. 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1869 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: DNA
US-09-261-006-1

Query Match 100.0%; Score 1869; DB 3; Length 1869;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1869; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANGACGGGATGATGGAACACAGACCTTTTCTAAGTTGGCTTACCTGGGCGACCGG 60
DB 1 ANGACGGGATGATGGAACACAGACCTTTTCTAAGTTGGCTTACCTGGGCGACCGG 60
QY 61 AGGATACGGGGAAGCTTAATCGGTACACCTGACGAGAGGCCAAGTGAAGGACACAG 120
DB 61 AGGATACGGGGAAGCTTAATCGGTACACCTGACGAGAGGCCAAGTGAAGGACACAG 120
QY 121 TACGAGAGACGAGTGTGTTGTAAGACCTTGAAGCGGCTCAAGGCGCTTCAATCGAAG 180
DB 121 TACGAGAGACGAGTGTGTTGTAAGACCTTGAAGCGGCTCAAGGCGCTTCAATCGAAG 180
QY 1861 GCGTCAATGCGGAGATTTCGCCACAGCGGACAGAAAGCTGCGCTTCACTGCTTTAAGAG 240

DB 181 GCGTCAATGCGGAGATTTCGCCACAGCGGACAGAAAGCTGCGCTTCACTGCTTTAAGAG 240
QY 241 GAGAGAGAGAGACCGAGATATGGTGGCGATATTCACAGACCTGAGCCCAAGAAATGTC 300
DB 241 GAGAGAGAGAGACCGAGATATGGTGGCGATATTCACAGACCTGAGCCCAAGAAATGTC 300
QY 301 CTCTCAATTAAGAAAGCTCCGCTCGATGACAGTGGAGACAGATTCAGAGAGACTTTAGT 360
DB 301 CTCTCAATTAAGAAAGCTCCGCTCGATGACAGTGGAGACAGATTCAGAGAGACTTTAGT 360
QY 361 GTGCGCTTCAAGAGAGAGAGATGAGAGACTTCGTCTTGAAGAGAGACTCCCGGTGAG 420
DB 361 GTGCGCTTCAAGAGAGAGAGATGAGAGACTTCGTCTTGAAGAGAGACTCCCGGTGAG 420
QY 421 TTGCAATATGGGATTTCTTGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 421 TTGCAATATGGGATTTCTTGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
QY 481 GCGGAGAGATATGAG 540
DB 481 GCGGAGAGATATGAG 540
QY 541 GATGCGATATGTTGTAAGCTCCCGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
DB 541 GATGCGATATGTTGTAAGCTCCCGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
QY 601 TACGACATATGCTTATGAG 660
DB 601 TACGACATATGCTTATGAG 660
QY 661 GAGGCGGTGATCCGAG 720
DB 661 GAGGCGGTGATCCGAG 720
QY 721 ATCAGGACACGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
DB 721 ATCAGGACACGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
QY 781 CCGCTGAGAGTGGGAG 840
DB 781 CCGCTGAGAGTGGGAG 840
QY 841 GCGGCGAGGATTAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
DB 841 GCGGCGAGGATTAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
QY 901 CACTGATTTACGGGCTTACGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTC 960
DB 901 CACTGATTTACGGGCTTACGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTC 960
QY 961 ACAGGATAGGAGAGTCTACCTTACGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTC 1020
DB 961 ACAGGATAGGAGAGTCTACCTTACGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTC 1020
QY 1021 GGGCGATATTCAGAGAGTCAAGACCTTCAGAGCGAGAGAGAGAGAGAGAGAGAGAGAG 1080
DB 1021 GGGCGATATTCAGAGAGTCAAGACCTTCAGAGCGAGAGAGAGAGAGAGAGAGAGAGAG 1080
QY 1081 GACCTGAGATAGAGAGGCTGTACCTCAGAGCGGAGGTTAAAGAGAGAGAGAGAGAGAG 1140
DB 1081 GACCTGAGATAGAGAGGCTGTACCTCAGAGCGGAGGTTAAAGAGAGAGAGAGAGAGAG 1140
QY 1141 ATAGTCTGTCACAGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1141 ATAGTCTGTCACAGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1201 CAGCTGATGCGAGAGAGAGGCTACTACTGCTCTCTGTAAGCCGCGGCGAGAGAGAG 1260
DB 1201 CAGCTGATGCGAGAGAGAGGCTACTACTGCTCTCTGTAAGCCGCGGCGAGAGAGAG 1260
QY 1261 TATAG 1320

QY	361	GTGCGCTTCAAGAGGAGGACGATGAGACCTTGCTTTTACGACGACGCTCCGGTCCG	420
Db	361	GTGCGCTTCAAGAGGAGGACGATGAGACCTTGCTTTTACGACGACGCTCCGGTCCG	420
QY	421	TTTCGACAAATATGCGAATCTTTTATGATGAGAGACGACGCTTCTGGTCTTTGAAC	480
Db	421	TTTCGACAAATATGCGAATCTTTTATGATGAGAGACGACGCTTCTGGTCTTTGAAC	480
QY	481	GGCGAGAGATATTCAGGAGCTCGAAGAACCGAGGTTTCGATGGGCCCTTGACAGGC	540
Db	481	GGCGAGAGATATTCAGGAGCTCGAAGAACCGAGGTTTCGATGGGCCCTTGACAGGC	540
QY	541	GATCGATATGTTGTGAACGTCGCCACCCGGAGGGAGCAGAGCCTCCCTGTTCAAGTTC	600
Db	541	GATCGATATGTTGTGAACGTCGCCACCCGGAGGGAGCAGAGCCTCCCTGTTCAAGTTC	600
QY	601	TACGACATATGCTTATGGAAGGACGGGGAGGAGACAAAGCTCTTCGAGAGGGCTCTTC	660
Db	601	TACGACATATGCTTATGGAAGGACGGGGAGGAGACAAAGCTCTTCGAGAGGGCTCTTC	660
QY	661	GAGCGGCTTGACTCCGACGAGAAAGAAATACTCTGAGGGCGAAGAAAAAAGCGGTTT	720
Db	661	GAGCGGCTTGACTCCGACGAGAAAGAAATACTCTGAGGGCGAAGAAAAAAGCGGTTT	720
QY	721	ATCAGCGACACGACCTGGCTGTACTCTCTGGAGCGCGAGCTTAAACGATCTACAGGGC	780
Db	721	ATCAGCGACACGACCTGGCTGTACTCTCTGGAGCGCGAGCTTAAACGATCTACAGGGC	780
QY	781	CCGCTCGAGCTGTGGGAAACCAAGCTCAGGAAAGAAAGTACTACTCTCCACTCCAGAT	840
Db	781	CCGCTCGAGCTGTGGGAAACCAAGCTCAGGAAAGAAAGTACTACTCTCCACTCCAGAT	840
QY	841	CGGGGACGGGTAAACCTCTGGCTCTGGAGCGGAAAGCCGAGCGTGTGTTACCCGGCAG	900
Db	841	CGGGGACGGGTAAACCTCTGGCTCTGGAGCGGAAAGCCGAGCGTGTGTTACCCGGCAG	900
QY	901	CACGTGAATTTACGGGCTTGACGACGAGCATATGGCAAAAGCTCTCTCATATGACACGCC	960
Db	901	CACGTGAATTTACGGGCTTGACGACGAGCATATGGCAAAAGCTCTCTCATATGACACGCC	960
QY	961	ACGAGGATAGCGGAGCTTACCTCTACGACGCGGAGCCTTAAACAGGTACCCAAATTAAC	1020
Db	961	ACGAGGATAGCGGAGCTTACCTCTACGACGCGGAGCCTTAAACAGGTACCCAAATTAAC	1020
QY	1021	GGGGCGATATTCAGGAAGCTCAAGACCTTCGAGCCGAGGACTTCCGCTTCAAGACAA	1080
Db	1021	GGGGCGATATTCAGGAAGCTCAAGACCTTCGAGCCGAGGACTTCCGCTTCAAGACAA	1080
QY	1081	GACCTCGAGATAGACGGCTGTATACCTCAGGCGGAGGTAAAGAGAGAAAGGCCCGCGTG	1140
Db	1081	GACCTCGAGATAGACGGCTGTATACCTCAGGCGGAGGTAAAGAGAGAAAGGCCCGCGTG	1140
QY	1141	ATAGCTTCCTTCACAGCGCGGGCCGAAAGGCGATGTACGACACCGCTTGCTTACGAGATG	1200
Db	1141	ATAGCTTCCTTCACAGCGCGGGCCGAAAGGCGATGTACGACACCGCTTGCTTACGAGATG	1200
QY	1201	CAGCTGATGCGGAGCAAGGCGCTACTACTGCTCTCTCGTGAACCCGCGGGGGAGGAGCGC	1260
Db	1201	CAGCTGATGCGGAGCAAGGCGCTACTACTGCTCTCTCGTGAACCCGCGGGGGAGGAGCGC	1260
QY	1261	TATAGCGAAGACTTCGCGCTCCGCGCTCGCTGAGAGAGACTGGCTTGAGAGACTTTGAGAG	1320
Db	1261	TATAGCGAAGACTTCGCGCTCCGCGCTCGCTGAGAGAGACTGGCTTGAGAGACTTTGAGAG	1320
QY	1321	ATAATGAAAGCGCATGAGAGAGTCTTCAAGCTCGAAACCGGAGCCGAGAGGAGGCGCTT	1380
Db	1321	ATAATGAAAGCGCATGAGAGAGTCTTCAAGCTCGAAACCGGAGCCGAGAGGAGGCGCTT	1380
QY	1381	GGAATTAACGGGCGATTAAGCTACGGCGGCTTCAATGACCAATGGGCCCTTACTCAAGGCGAC	1440
Db	1381	GGAATTAACGGGCGATTAAGCTACGGCGGCTTCAATGACCAATGGGCCCTTACTCAAGGCGAC	1440
QY	1441	CTCTTCAAAGCAGGAATTAAGCAGAGAACGGCGATTAAGCTACTGGCTTACCACGCTACGCCCTTC	1500

Db	1441	CTTTCAGGAGGAATTAAGCGAGAACGCATTAAGTACTGGCTCACAGCTAAGCCCTTC	1500
QY	1501	TCGGACATATAGGGCTCTGGTACGACGTGAGGTCAATCGGGCCCAATCCGTTAGAGAACGAG	1560
Db	1501	TCGGACATATAGGGCTCTGGTACGACGTGAGGTCAATCGGGCCCAATCCGTTAGAGAACGAG	1560
QY	1561	AACCTCAGGAAGCTCAGACCCCGCTGTTCTTAAGCTCAGAACGTAAGGGGCCGATACCTCTTA	1620
Db	1561	AACCTCAGGAAGCTCAGACCCCGCTGTTCTTAAGCTCAGAACGTAAGGGGCCGATACCTCTTA	1620
QY	1621	ATCCATCTCGCTTGAGAGCTACCGCTGTCCGCTGCACACAGACCTTATGTTCTTAAACGTG	1680
Db	1621	ATCCATCTCGCTTGAGAGCTACCGCTGTCCGCTGCACACAGACCTTATGTTCTTAAACGTG	1680
QY	1681	CTCAAGGACATATGGGCAAGAGACCTACATAGCGATATTCAAGCGCGCGCCACAGGCGCAC	1740
Db	1681	CTCAAGGACATATGGGCAAGAGACCTACATAGCGATATTCAAGCGCGCGCCACAGGCGCAC	1740
QY	1741	AGGCTCCGGGAGAACCCCGAGGACACAGGCCGGAAGCGCTTACAGGCTCTTCATAGATCTTTC	1800
Db	1741	AGGCTCCGGGAGAACCCCGAGGACACAGGCCGGAAGCGCTTACATAGATCTTTC	1800
QY	1801	GAGCGCAAGCTCAAGAAAGTACGAGAGAGGGCTTTGAGTACAGAAAGATATCTCAAGGGCAAT	1860
Db	1801	GAGCGCAAGCTCAAGAAAGTACGAGAGAGGGCTTTGAGTACAGAAAGATATCTCAAGGGCAAT	1860
QY	1861	GGGAAGTGA 1869	
Db	1861	GGGAAGTGA 1869	

RESULT 6
 US-09-609-570-1
 ; Sequence 1, Application US/09609570
 ; Patent No. 6465204
 ;
 GENERAL INFORMATION:
 ;
 APPLICANT: Murphy et al.
 TITLE OF INVENTION: Amidases
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CAPELLA, BYRNE, BAIN, GILFILLAN
 CECCHI, STEWART & OLSTEIN
 STREET: 6 BECKER FARM ROAD
 CITY: ROSELAND
 STATE: NEW JERSEY
 COUNTRY: USA
 ZIP: 07068
 ;
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5 INCH DISKETTE
 COMPUTER: IBM PS/2
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: WORD PERFECT 5.1
 ;
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/609,570
 FILING DATE: 30-Jun-2000
 CLASSIFICATION: <Unknown>
 ;
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/427,372
 FILING DATE: <Unknown>
 ;
 ATTORNEY/AGENT INFORMATION:
 NAME: Charles J. Herron
 REGISTRATION NUMBER: 28,019
 REFERENCE/DOCKET NUMBER: 331400-53
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-994-1700
 TELEFAX: 201-994-1744
 ;
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1869 NUCLEOTIDES
 TYPE: NUCLEIC ACID
 STRANDEDNESS: SINGLE
 TOPOLOGY: LINEAR

MOLECULE TYPE: DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-609-570-1

Query Match 100.0%; Score 1869; DB 4; Length 1869;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1869; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGACGGCATTCGAATGGAACCAACGAGACCTTTTCTAAGTTCGCTTACCTGGGCGACCCG 60
DB 1 ATGACGGCATTCGAATGGAACCAACGAGACCTTTTCTAAGTTCGCTTACCTGGGCGACCCG 60
QY 61 AGGATACGGGGAACCTTAATCGGCTACACCTGACGAAAGGCGCAACATGAAAGCAACAAG 120
DB 61 AGGATACGGGGAACCTTAATCGGCTACACCTGACGAAAGGCGCAACATGAAAGCAACAAG 120
QY 121 TACGAGACACGGCTTGTGTGTAAGACCTTGAACGAGGCTCAACAGGCGCTTCAACGAAAC 180
DB 121 TACGAGACACGGCTTGTGTGTAAGACCTTGAACGAGGCTCAACAGGCGCTTCAACGAAAC 180
QY 181 GCCTCAATGCCGAGATTTCCGACAGGCGCAAGAGCTGCTTCACTGCTTTAAAGAG 240
DB 181 GCCTCAATGCCGAGATTTCCGACAGGCGCAAGAGCTGCTTCACTGCTTTAAAGAG 240
QY 241 GAGAAAGAGAGACCGAGATATGGGTGCGGATATTCAGACCCCTGAGCGCCCAAGAAAGTC 300
DB 241 GAGAAAGAGAGACCGAGATATGGGTGCGGATATTCAGACCCCTGAGCGCCCAAGAAAGTC 300
QY 301 CTCTCAACTTAAACCTCCGCTCGATGACAGTGAAGAGCAAGTTCAGAGACCTTAAAGTT 360
DB 301 CTCTCAACTTAAACCTCCGCTCGATGACAGTGAAGAGCAAGTTCAGAGACCTTAAAGTT 360
QY 361 GTGCGCTCAAGAGAGGAGGAGATGAGGAGCTTGTGTAAGACAGACGCTCCGCGCTTG 420
DB 361 GTGCGCTCAAGAGAGGAGGAGATGAGGAGCTTGTGTAAGACAGACGCTCCGCGCTTG 420
QY 421 TTCGACAAATATGAGGATCTTGTATGAGAGAGAGACGCTTGTGTAAGAGAGAG 480
DB 421 TTCGACAAATATGAGGATCTTGTATGAGAGAGAGACGCTTGTGTAAGAGAGAG 480
QY 481 GCCGAGAGATATTCAGACAGTTCGAGAGAGCCGAGTTCGAGTGGCTTGTGCAAGGCG 540
DB 481 GCCGAGAGATATTCAGACAGTTCGAGAGAGCCGAGTTCGAGTGGCTTGTGCAAGGCG 540
QY 541 GATGCGATAGTGTGAAGCTCCGCGACCGGAGGAGGAGCAAGCTTGTGTAAGAGAG 600
DB 541 GATGCGATAGTGTGAAGCTCCGCGACCGGAGGAGGAGCAAGCTTGTGTAAGAGAG 600
QY 601 TACGACATAGTCTATGGAAGAGAGGAGGAGAGAGAGCTTGTGTAAGAGAGAG 660
DB 601 TACGACATAGTCTATGGAAGAGAGGAGGAGAGAGAGAGCTTGTGTAAGAGAGAG 660
QY 661 GAGGCGGTGACCTCCGACGAAAGAGAAATCTCTGAGGCGCAAGAAAAAAGCGGTTTC 720
DB 661 GAGGCGGTGACCTCCGACGAAAGAGAAATCTCTGAGGCGCAAGAAAAAAGCGGTTTC 720
QY 721 ATCAGGAGACAGACAGTGTGATGCTCTGAGGAGGAGGAGGAGTAAACGATCTAGAGG 780
DB 721 ATCAGGAGACAGACAGTGTGATGCTCTGAGGAGGAGGAGGAGTAAACGATCTAGAGG 780
QY 781 CCGCTGAGAGCTGCGGAGAGCAAGCTACAGGAAAGAGTCTAATCTCCACTCAAGAT 840
DB 781 CCGCTGAGAGCTGCGGAGAGCAAGCTACAGGAAAGAGTCTAATCTCCACTCAAGAT 840
QY 841 GCGGCGAGGGAACCTTGGCTCTGGGAGGGAAGCGAGGCTGTGTTACCGGCGAC 900
DB 841 GCGGCGAGGGAACCTTGGCTCTGGGAGGGAAGCGAGGCTGTGTTACCGGCGAC 900
QY 901 CACTGATTTTACGGGCTTGAAGCTCAGCGATGCAAGAGCATGCTCTCATCATGACCGCC 960
DB 901 CACTGATTTTACGGGCTTGAAGCTCAGCGATGCAAGAGCATGCTCTCATCATGACCGCC 960
QY 961 ACAGAGATAGGCGAGCTTACTCTTACAGACGCGGAGCTGAACAGGTCAACGAAATACAC 1020
```

```
DB 961 ACAGAGATAGGCGAGCTTACTCTTACAGACGCGGAGCTGAACAGGTCAACGAAATACAC 1020
QY 1021 GGGCGCATATTTAGGAAGCTCAAGACCTTGAAGCCGAGGCACTTCCGTTCAAGAGCAAA 1080
DB 1021 GGGCGCATATTTAGGAAGCTCAAGACCTTGAAGCCGAGGCACTTCCGTTCAAGAGCAAA 1080
QY 1081 GACCTGAGATATAGCGGCTGTGATCTCAGCGCGGAGGTTAAAGAGAGAGCCCGCGTG 1140
DB 1081 GACCTGAGATATAGCGGCTGTGATCTCAGCGCGGAGGTTAAAGAGAGAGCCCGCGTG 1140
QY 1141 ATAGCTTGTCCAGGCGGCGGAGAGGAGCATATGATAGGACACCGCTGTGTAAGAGAG 1200
DB 1141 ATAGCTTGTCCAGGCGGCGGAGAGGAGCATATGATAGGACACCGCTGTGTAAGAGAG 1200
QY 1201 CAGCTGATGGCGAGCAAGGAGCTTACTGCTGCTGCTGTAACCCGCGCGGAGCGAGCG 1260
DB 1201 CAGCTGATGGCGAGCAAGGAGCTTACTGCTGCTGCTGTAACCCGCGCGGAGCGAGCG 1260
QY 1261 TATAGCGAAGACTTCCGCGCTCCGCTCTGAGAGAGAGCTGCTTGAAGAGACTTTAGAG 1320
DB 1261 TATAGCGAAGACTTCCGCGCTCCGCTCTGAGAGAGAGCTGCTTGAAGAGACTTTAGAG 1320
QY 1321 ATATGACGCGCATCGAGAGATTTCTCAAGCTCGAACCAGCGAGCGAGAGAGCGCTT 1380
DB 1321 ATATGACGCGCATCGAGAGATTTCTCAAGCTCGAACCAGCGAGCGAGAGAGCGCTT 1380
QY 1381 GGAATAGCGGCGATTAAGCTACGCGGCTTCAATGACCACTGAGGCTTGAAGAGCGAG 1440
DB 1381 GGAATAGCGGCGATTAAGCTACGCGGCTTCAATGACCACTGAGGCTTGAAGAGCGAG 1440
QY 1441 CTCTCAAGGAGGATTAAGCGGAGAGGAGCATATGATGCTGCTACAGCTTACGCGCTTC 1500
DB 1441 CTCTCAAGGAGGATTAAGCGGAGAGGAGCATATGATGCTGCTACAGCTTACGCGCTTC 1500
QY 1501 TCGGACATAGGCGCTGTGATACGAGCTGAGAGGATCATGCGCAATCCGTTAGAGAGAG 1560
DB 1501 TCGGACATAGGCGCTGTGATACGAGCTGAGAGGATCATGCGCAATCCGTTAGAGAGAG 1560
QY 1561 AACTTACAGAGAGCTACGCGCGCTGTCTACGCTCAGAAAGTGAAAGGCGCGGATCTCTA 1620
DB 1561 AACTTACAGAGAGCTACGCGCGCTGTCTACGCTCAGAAAGTGAAAGGCGCGGATCTCTA 1620
QY 1621 ATCCACTGCTTATGAGACTACGCGCTGCTGCGTACGACAGAGCTTATGTTTACAAAGCT 1680
DB 1621 ATCCACTGCTTATGAGACTACGCGCTGCTGCGTACGACAGAGCTTATGTTTACAAAGCT 1680
QY 1681 CTCAGAGACATGGGCAAGGAGGCTTACATAGCGATATTCAGCGCGCGCCCAAGCGGCG 1740
DB 1681 CTCAGAGACATGGGCAAGGAGGCTTACATAGCGATATTCAGCGCGCGCCCAAGCGGCG 1740
QY 1741 AGCGTCCGCGGAAGCCCGAGGAGCAAGGCGGAGGCTTACAGGCTTCTTACAGTCTTC 1800
DB 1741 AGCGTCCGCGGAAGCCCGAGGAGCAAGGCGGAGGCTTACAGGCTTCTTACAGTCTTC 1800
QY 1801 GAGCGCAACCTCAAGAGATACGAGAGGAGGCTTGAAGTGAAGAGATCTCAAGGAGGAA 1860
DB 1801 GAGCGCAACCTCAAGAGATACGAGAGGAGGCTTGAAGTGAAGAGATCTCAAGGAGGAA 1860
QY 1861 GGGAGCTGA 1869
DB 1861 GGGAGCTGA 1869
```

RESULT 7
US-09-016-080-2
; Sequence 2, Application US/09016080
; Patent No. 6133012
; GENERAL INFORMATION:
; APPLICANT: Ishikawa, Kazuhiko
; APPLICANT: Matsui, Ikuro
; APPLICANT: Ishida, Hiroyasu
; APPLICANT: Kosugi, Yoshisugu


```

; APPLICANT: Higauchi, Katsuniko
; TITLE OF INVENTION: THERMOSTABLE ACYL PEPTIDE HYDROLASE AND GENE ENCODING
; FILE OF INVENTION: THE SAME
; FILE REFERENCE: 07898/022001
; CURRENT APPLICATION NUMBER: US/09/016,080
; CURRENT FILING DATE: 1998-01-30
; EARLIER APPLICATION NUMBER: JAPAN 18381/1997
; EARLIER FILING DATE: 1997-01-31
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1896
; TYPE: DNA
; ORGANISM: Pyrococcus horikoshii
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1896)
US-09-016-080-2

```

```

Query Match          6.4%; Score 120.4; DB 3; Length 1896;
Best Local Similarity 50.3%; Pred. No. 5.2e-21;
Matches 351; Conservative 0; Mismatches 341; Indels 6; Gaps 2;

```

```

QY 1120 AAAGAGGAGAGAGCCCGGTATGCTTCCTCCACGGCGCGGAGAGGCGATGACGA 1179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1198 AAAGGAAAGAGTCCAGCTATCTAGAGTCCAGGTGCTCTAAACCGCTTACGGT 1257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1180 CACCGCTTCGTACGAGATGACAGCTGATGCGAGCAAGAGCTACTGCTGCTGTG 1239
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1258 TACGCTTTTATGACGAGGTTCACGTTTAACTCTAAAGGCTTCGTGATTTCTCA 1317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1240 AACCGCGCGGACGCGGCTATAGGAGAGACTTCGCCGCTCCCGCTCTGAGAGACT 1299
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1318 AATCTAGAGAGGAGGAGTGTCTACGAGAGAGTTCGCGGATA--TAAAGGACACTAT 1374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1300 GCGTTGAGGACTTTGAGACATATATGACGATGAGGATGCTTCAAGCTCGAACCG 1359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1375 GGGAGAGAGGATTCACGAGTTTATGAGGTATGCTCATGAGCATTAAGAGATTGAC 1434
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1360 CAGGCGACAGAGGAGCGGCTTGAATTAACGGGCTAAAGCTACGGCGCTTCAAGACCAC 1419
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1435 TTGATGATGAGGAGGAGGAGGCTACGAGGAGTTCATGAGTGGCTTCATGACGAC 1494
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1420 TGGGCTTGACTGAGAGCGACTCTTCAAGGCGAGGATTAAGCGGAAACGGCATTAAGCTAC 1479
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1495 TGGATAGTCCGACATACCAAGGTTAAAGCGCTGTAAACCGAGATCAATTTCAAT 1554
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1480 TGGCTCACACGACTACGCTTCTCGACATAGGGCTGTGTACGAGCTGAGGTGATCGG 1539
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1555 TGGATTAAGCTTCTCGGAGCAACGAGATATAGTTTACTTGTCTCCAGATCAATTAAGA 1614
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1540 CCAAAATCCGTTAAGAAC--GAGAACTTACGAGCTACGCCGCTGTCTACGCTCAG 1596
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1615 AAGATATCCCTGGAGCAACTTGAAGGTTATTTGGGAAAAAGAGCCATTAAGTACGGCTCC 1674
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1597 AAGCTAAGGCGCGGATCTCTAATCCACTCGCTTGAAGACATACCGCTCTCCGCTCGAC 1656
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1675 AAGCTGAAGAACTCCCTGCTTAAATCACTTACCGAAGACTACAGGTTCGCTTCC 1734
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1657 CAGAGCTTATGTTTACACAGTGTCTACAGGATGAGGCGCAAGGCTCATAGGAGAT 1716
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1735 GAGGCAATGCAACTTTCATATCCCTAAATATCTAGGAGAGAGATGATTAATTGGCAATA 1794
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1717 TTCAGAGCGGCGCCGACGCGCACAGCTCCGCGAAGCCGAGGACAGGCGGAGCGG 1776
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1795 TTCCGAGGAGAAATATGATGACCTAAGTAGTGTGGAGGCAAGACAGGCGTTAAAGA 1854
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1777 TACAGGCTCTTCAATAGATTCTCTGAGCGGACGCAAGCTCAA 1814
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1855 CTTGAACATAATGACGATGATGAGAGAAATGGCTTAA 1892
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 8

```

US-08-211-682-24
; Sequence 24, Application US/08211682
; Patent No. 5670333
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT:
; TITLE OF INVENTION: EXPRESSION OF POLYPEPTIDES IN E. COLI UNDER
; TITLE OF INVENTION: CONTROL OF THE E. COLI MDH-GENE PROMOTER
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/211,682
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1598 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 286..1269
; FEATURE:
; NAME/KEY: -10_signal
; LOCATION: 228..233
; FEATURE:
; NAME/KEY: -35_signal
; LOCATION: 205..210
US-08-211-682-24

```

```

Query Match          2.6%; Score 48.2; DB 1; Length 1598;
Best Local Similarity 46.1%; Pred. No. 0.0065;
Matches 161; Conservative 0; Mismatches 188; Indels 0; Gaps 0;

```

```

QY 1505 ACATAGGCGCTCTGTACGAGCTGAGGTGATCGGAGCCCAATCCGTTGAGAGAGAGACT 1564
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 329 AGATCGGCTACAGCTCTCTTCCGATCCGCCGCGGAGAGATGCTGGCAAGGACG 388
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1565 TCAGGAAGCTACAGCCGCTTCTTACGCTCAGAACGTGAAGGCGGCGATCTTAATCC 1624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 389 CGGTGATCTCCAGCACTTTTGAAGATCCCCAGGCGCATGAAGGCCCTGAGAGGGGTGTCA 448
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1625 ACTGCGTTGAGGACTACCGGTCTCCGTGACCAAGAGCTTATGTTCTACAGTGTCTCA 1684
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 449 TGGAGCTGGAGAGACTGCGCTTCCCTGCTTCCGCGGCGCTCGAGGCCACGACGCCCA 508
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1685 AGGACATGAGGCAAGAGCTTACATAGCGATTTTCAAGCGCGGCGCCACGCGCACAGCG 1744
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 509 AGGTGCGCTTCAAGAGAGCGCGCATACGCTCTCTGCTGAGGCGCGGCGCGCGCAAGGCG 568
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1745 TCCGCGAAGCCCGAGGACACAGCGCAAGCGCTACAGGCTTCTCATAGATTTCTTGAGC 1804
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 569 GGATGAGAGCGCCGCGGACTTTTGCAGGTAAACGCAAGATCTTACCGAGAGAGGCGCGG 628
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1805 GCAAGCTCAAGATGACGAGGAGGCGCTTGAAGTAAAGAAATCTCAA 1853
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 629 CCTGGCGGAGGTGGCCAGAGAGCGTCAAGGTGTGTGTGGGCAA 677
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 9

```

US-09-024-020B-1
; Sequence 1, Application US/09024020B
; Patent No. 6030810
; GENERAL INFORMATION:
; APPLICANT: DELGADO, STEPHEN G.
; APPLICANT: DIETRICH, PAUL S.
; APPLICANT: FISH, LINDA M.
; APPLICANT: HERMAN, RONALD C.
; APPLICANT: SANGAMESWARAN, LAKSHMI

```



```

; TITLE OF INVENTION: NOVEL CLONED TETRODOTOXIN-SENSITIVE
; TITLE OF INVENTION: SODIUM CHANNEL 1-SUBUNIT AND A SPLICE VARIANT THEREOF
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JANET PAULINE CLARK
; STREET: 3401 HILLVIEW AVENUE, MS A2-250
; CITY: PALO ALTO
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94304-1397
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/024,020B
; FILING DATE: 16-FEB-1998
; CLASSIFICATION: 536
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/039,447
; FILING DATE: 26-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: CLARK, JANET P.
; REGISTRATION NUMBER: 34,799
; REFERENCE/DOCKET NUMBER: R0020B-REG
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 852-3097
; TELEFAX: (650) 855-5322
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-024-020B-1

Query Match
Best Local Similarity 2.5%; Score 46.4; DB 3; Length 5977;
Matches 131; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 1200 GCACGTGATGGCGAGGAGGCTACTACTGCTCTTCGTAACCCGCGCGGACGCGAGG 1259
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4146 GAAGCTCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4205
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1260 CTATACGGAAGACTTGGCGCTCGCGCTCTCGAGAGGAGGAGGAGGAGGAGGAGGAGG 1319
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4206 CAATGTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4265
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1320 CATTAATGAACGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1379
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4266 CATCATGTAATGCGGCTGTAGATTCGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4325
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1380 TGGAAATGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1439
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4326 CATCTACATGTAATCTACTCTGTCATCTTCATCATCTTCGCTCTTCTTCAACCTCA 4385
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1440 CCTCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1471
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4386 CCTGTTCATCGGTGTCTCATCTCATGCAACTTCA 4417
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
US-09-425-043-1
; Sequence 1, Application US/09425043
; Patent No. 6335172
; GENERAL INFORMATION:
; APPLICANT: DELGADO, STEPHEN G.
; APPLICANT: DIETRICH, PAUL S.
; APPLICANT: FISH, LINDA M.
; APPLICANT: HERMAN, RONALD C.
; APPLICANT: SANGAMESWARAN, LAKSHMI

```

```

; TITLE OF INVENTION: NOVEL CLONED TETRODOTOXIN-SENSITIVE
; TITLE OF INVENTION: SODIUM CHANNEL 1-SUBUNIT AND A SPLICE VARIANT THEREOF
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JANET PAULINE CLARK
; STREET: 3401 HILLVIEW AVENUE, MS A2-250
; CITY: PALO ALTO
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94304-1397
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/425,043
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 09/024,020
; FILING DATE: 16-FEB-1998
; APPLICATION NUMBER: US 60/039,447
; FILING DATE: 26-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: CLARK, JANET P.
; REGISTRATION NUMBER: 34,799
; REFERENCE/DOCKET NUMBER: R0020B-REG
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 852-3097
; TELEFAX: (650) 855-5322
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-425-043-1

Query Match
Best Local Similarity 2.5%; Score 46.4; DB 4; Length 5977;
Matches 131; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 1200 GCACGTGATGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1259
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4146 GAAGCTCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4205
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1260 CTATACGGAAGACTTGGCGCTCGCGCTCTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1319
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4206 CAATGTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4265
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1320 CATTAATGAACGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1379
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4266 CATCATGTAATGCGGCTGTAGATTCGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4325
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1380 TGGAAATGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1439
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4326 CATCTACATGTAATCTACTCTGTCATCTTCATCATCTTCGCTCTTCTTCAACCTCA 4385
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1440 CCTCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1471
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4386 CCTGTTCATCGGTGTCTCATCTCATGCAACTTCA 4417
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
US-09-024-020B-2
; Sequence 2, Application US/09024020B
; Patent No. 6030810
; GENERAL INFORMATION:
; APPLICANT: DELGADO, STEPHEN G.
; APPLICANT: DIETRICH, PAUL S.
; APPLICANT: FISH, LINDA M.

```


APPLICANT: HERMAN, RONALD C.
APPLICANT: SANGMESMARAN, LAKSHMI
TITLE OF INVENTION: NOVEL CLONED TETRODOPOXIN-SENSITIVE
TITLE OF INVENTION: SODIUM CHANNEL I-SUBUNIT AND A SPLICE VARIANT THEREOF
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: JANET PAULINE CLARK
STREET: 3401 HILLYVIEW AVENUE, MS A2-250
CITY: PALO ALTO
STATE: CA
COUNTRY: U.S.A.
ZIP: 94304-1397
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/024,020B
FILING DATE: 16-FEB-1998
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/039,447
FILING DATE: 26-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: CLARK, JANET P.
REGISTRATION NUMBER: 34,799
REFERENCE/DOCKET NUMBER: R0020B-REG
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 852-3097
TELEFAX: (650) 855-5322
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 6007 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-024-020B-2

Query Match 2.5%; Score 46.4; DB 3; Length 6007;
Best Local Similarity 48.2%; Pred. No. 0.029;
Matches 131; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 1200 GCAGCTGATGGCGAGCAAGGCGTACTACTGCTGCTGTAACCCGCGCGGCGAGCG 1259
DB 4176 GAAGCTCATGAGGAGGCAACAGCAGGATCCGATGGAGAAGATGTCAGATCAACTTTGA 4235
QY 1260 CTATTAGCGAAGACTTGCCTCGCGCTCGAGAGAGACTGGCTTGAGAGACTTTGAGGA 1319
DB 4236 CATGTGCGAGCGAGGACTGCTGCTTTCTTCAAGTGGCAACCTTCAAAAGCTGGATGA 4295
QY 1320 CATATGAAAGCGCATGAGAGATTCTTCAAGCTGGAACCGCAGCGCAGAGAGCGGT 1379
DB 4296 CATCATGTATGCGGTGTAGATTCCGAAAGCCAGAGCAGACGCTGACTACAGAGGCAA 4355
QY 1380 TGGAAATTAAGCGGATAGCTAAGCGCGGCTTCAATGACCACTGGCGCTTGAAGAGCGA 1439
DB 4356 CATCTACATGTATATCTACTTCTCATCTTATCATCTTGGCTCTTCAACCTTCAA 4415
QY 1440 CCTCTTCAAGGAGCAATTAAGCGAGAACGGCA 1471
DB 4416 CCTGTTCATCGGTGTCTATCATGCAACTTCA 4447

RESULT 12
US-09-425-043-2
; Sequence 2, Application US/09425043
; Patent No. 6335172
; GENERAL INFORMATION:
; APPLICANT: DELGADO, STEPHEN G.
; APPLICANT: DIETRICH, PAUL S.
; APPLICANT: FISH, LINDA M.

APPLICANT: HERMAN, RONALD C.
APPLICANT: SANGMESMARAN, LAKSHMI
TITLE OF INVENTION: NOVEL CLONED TETRODOPOXIN-SENSITIVE
TITLE OF INVENTION: SODIUM CHANNEL I-SUBUNIT AND A SPLICE VARIANT THEREOF
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: JANET PAULINE CLARK
STREET: 3401 HILLYVIEW AVENUE, MS A2-250
CITY: PALO ALTO
STATE: CA
COUNTRY: U.S.A.
ZIP: 94304-1397
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/425,043
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/024,020
FILING DATE: 16-FEB-1998
APPLICATION NUMBER: US 60/039,447
FILING DATE: 26-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: CLARK, JANET P.
REGISTRATION NUMBER: 34,799
REFERENCE/DOCKET NUMBER: R0020B-REG
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 852-3097
TELEFAX: (650) 855-5322
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 6007 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-425-043-2

Query Match 2.5%; Score 46.4; DB 4; Length 6007;
Best Local Similarity 48.2%; Pred. No. 0.029;
Matches 131; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 1200 GCAGCTGATGGCGAGCAAGGCGTACTACTGCTGCTGTAACCCGCGCGGCGAGCG 1259
DB 4176 GAAGCTCATGAGGAGGCAACAGCAGGATCCGATGGAGAAGATGTCAGATCAACTTTGA 4235
QY 1260 CTATTAGCGAAGACTTGCCTCGCGCTCGAGAGAGACTGGCTTGAGAGACTTTGAGGA 1319
DB 4236 CATGTGCGAGCGAGGACTGCTGCTTTCTTCAAGTGGCAACCTTCAAAAGCTGGATGA 4295
QY 1320 CATATGAAAGCGCATGAGAGATTCTTCAAGCTGGAACCGCAGCGCAGAGAGCGGT 1379
DB 4296 CATCATGTATGCGGTGTAGATTCCGAAAGCCAGAGCAGACGCTGACTACAGAGGCAA 4355
QY 1380 TGGAAATTAAGCGGATAGCTAAGCGCGGCTTCAATGACCACTGGCGCTTGAAGAGCGA 1439
DB 4356 CATCTACATGTATATCTACTTCTCATCTTATCATCTTGGCTCTTCAACCTTCAA 4415
QY 1440 CCTCTTCAAGGAGCAATTAAGCGAGAACGGCA 1471
DB 4416 CCTGTTCATCGGTGTCTATCATGCAACTTCA 4447

RESULT 13
US-09-024-020B-7
; Sequence 7, Application US/09024020B
; Patent No. 6030810
; GENERAL INFORMATION:
; APPLICANT: DELGADO, STEPHEN G.


```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```


XX WPI: 1998-063135/06.
DR P-PSDB; AAM41248.
XX Nucleic acid encoding heat-stable amidase from *Thermococcus* GUSL5 -
PT useful for removing N-terminal amino acids from synthetic peptide(s)
PT and peptidomimetic(s)
XX
PS Claim 2; Page 39-41; 56bp; English.
XX This novel isolated nucleic acid encodes a 622 amino acid
CC thermostable amidase (see AAM41248) of the archaeobacterium
CC *Thermococcus* GUSL5. It was isolated from a genomic library of
CC GUSL5 by screening for amidase activity, and subjecting DNA from a
CC positive clone to PCR amplification (see AAM2888-89). The nucleic
CC acid can be used in the production of amidase in host cells, and to
CC identify related sequences encoding similar enzymes. The encoded
CC amidase is used to remove Arg, Phe or Met from the N-terminus of
CC synthetic peptides or peptidomimetics (claimed). Removal of the
CC N-terminal residue can be done even in the presence of a more
CC reactive ester bond (very difficult to achieve non-enzymatically).
CC The amidase is selective for L-amino acids and can therefore be
CC used to produce optically active compounds. The protein tolerates
CC temperatures up to at least 70 degC and high concentrations of
CC organic solvent (e.g. over 40% dimethyl sulphoxide) so can cleave
CC bonds that are normally resistant.
XX
SO Sequence 1869 BP; 476 A; 486 C; 567 G; 340 T; 0 other;
Query Match 100.0%; Score 1869; DB 19; Length 1869;
Best Local Similarity 100.0%; Pred. No.: 0;
Matches 1869; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ATGACGGGATCGAATGGAACACGAGACCTTTCTAAGTCCGCTACCGGGCGACCG 60
DB 1 ATGACGGGATCGAATGGAACACGAGACCTTTCTAAGTCCGCTACCGGGCGACCG 60
OY 61 AGGATACGGGGAATTAATCGCGTACACCTTCAGCAAGGCAACATGAAAGCAACAG 120
DB 61 AGGATACGGGGAATTAATCGCGTACACCTTCAGCAAGGCAACATGAAAGCAACAG 120
OY 121 TAGAAGACGCGGTGTGTGTTGAAGACCTTGAACGGGCTCAAGGGCTTCATCGAGAC 180
DB 121 TAGAAGACGCGGTGTGTGTTGAAGACCTTGAACGGGCTCAAGGGCTTCATCGAGAC 180
OY 121 TAGAAGACGCGGTGTGTGTTGAAGACCTTGAACGGGCTCAAGGGCTTCATCGAGAC 180
DB 121 TAGAAGACGCGGTGTGTGTTGAAGACCTTGAACGGGCTCAAGGGCTTCATCGAGAC 180
OY 181 GCCTCAATGCCGAGGATTTGGCAGACGGCAGAAAGTCGCTTCACCTTCTTAACGAG 240
DB 181 GCCTCAATGCCGAGGATTTGGCAGACGGCAGAAAGTCGCTTCACCTTCTTAACGAG 240
OY 241 GAGAGAGAGAGACCGAGATATGGTGGCCGATATCCAGACCTTGAGCGCCAAAGATC 300
DB 241 GAGAGAGAGAGACCGAGATATGGTGGCCGATATCCAGACCTTGAGCGCCAAAGATC 300
OY 301 CTCTCAACTAAACACGTCGCTCGATGACATGGAAGACGATTCAGAGAACTCTTAGT 360
DB 301 CTCTCAACTAAACACGTCGCTCGATGACATGGAAGACGATTCAGAGAACTCTTAGT 360
OY 361 GTGGGCTTCAAGAGAGAGAGATGAGAGACTTGTCTTTGACAGCAGCGTCCGGTGG 420
DB 361 GTGGGCTTCAAGAGAGAGAGATGAGAGACTTGTCTTTGACAGCAGCGTCCGGTGG 420
OY 421 TTCGACATATATGGGATTTCTTGATGAGAGAGACGAGCTTGGGTTCTTGACATGAG 480
DB 421 TTCGACATATATGGGATTTCTTGATGAGAGAGACGAGCTTGGGTTCTTGACATGAG 480
OY 481 GCCGAGAGATATATCGAGAGCTTCGAGAAACCGGAGTTTTCGAGTGGCTTCGACAGGC 540
DB 481 GCCGAGAGATATATCGAGAGCTTCGAGAAACCGGAGTTTTCGAGTGGCTTCGACAGGC 540
OY 541 GATGCGATAGTTGTGAACGTCGCCGACCGGAGGGAGCAAGCCTGCCCTGTTCAAGTTC 600
DB 541 GATGCGATAGTTGTGAACGTCGCCGACCGGAGGGAGCAAGCCTGCCCTGTTCAAGTTC 600

OY 601 TAGACATAGTCTCTATGGAAGACGGGGAGGAAGAAAGCTCTTCGAGAGGGTCTTC 660
DB 601 TAGACATAGTCTCTATGGAAGACGGGGAGGAAGAAAGCTCTTCGAGAGGGTCTTC 660
OY 661 GAGCGGTTGACTCCGACGGAAGAGAAATCTCTGAGGGGCAAGAAAAAGCGGTTTC 720
DB 661 GAGCGGTTGACTCCGACGGAAGAGAAATCTCTGAGGGGCAAGAAAAAGCGGTTTC 720
OY 721 ATCAGGAGACGACATGCTGTGTACCTCTGTGGGACGGGAGCTTAAACCATCTACAGGGC 780
DB 721 ATCAGGAGACGACATGCTGTGTACCTCTGTGGGACGGGAGCTTAAACCATCTACAGGGC 780
OY 781 CCCTGACGCTCTGGGAGGCAAGCTCAGAGAAAGGTTACTTCTCTCACTCCAGAT 840
DB 781 CCCTGACGCTCTGGGAGGCAAGCTCAGAGAAAGGTTACTTCTCTCACTCCAGAT 840
OY 841 GCGGGAGGGTAAACCTGTGCTGTGGAGCGGAGCCGAGCGTGTACCGGCGAC 900
DB 841 GCGGGAGGGTAAACCTGTGCTGTGGAGCGGAGCCGAGCGTGTGTACCGGCGAC 900
OY 901 CACTGATTTACGGGCTGTGACGTCACGATGGCAAGCATGTGCTCTCATGACCGCC 960
DB 901 CACTGATTTACGGGCTGTGACGTCACGATGGCAAGCATGTGCTCTCATGACCGCC 960
OY 961 ACAGGATAGCGAGCTCTACCTCTACGACGCGGACGCTGAACAGGTCACCGAATCAAC 1020
DB 961 ACAGGATAGCGAGCTCTACCTCTACGACGCGGACGCTGAACAGGTCACCGAATCAAC 1020
OY 1021 GGGCCGATTTACGGAAGACTCAAGACCTTCAGCCGACGCTTCCTCAAGAGCAAA 1080
DB 1021 GGGCCGATTTACGGAAGACTCAAGACCTTCAGCCGACGCTTCCTCAAGAGCAAA 1080
OY 1081 GACCTGAGATAGACGGCTGTGTACCTACGCGGAGGTTAAAGAGAGAAAGCCCGGG 1140
DB 1081 GACCTGAGATAGACGGCTGTGTACCTACGCGGAGGTTAAAGAGAGAAAGCCCGGG 1140
OY 1141 ATAGTCTCTCCACGCGGGGCGGAGGAGGATACGAGACCGCTTCGTACAGAGAT 1200
DB 1141 ATAGTCTCTCCACGCGGGGCGGAGGAGGATACGAGACCGCTTCGTACAGAGAT 1200
OY 1201 CACTGATGGCGAGCAAGGGCTACTACTGTGCTGTGTAACCGGGCGGACGCGGCG 1260
DB 1201 CACTGATGGCGAGCAAGGGCTACTACTGTGCTGTGTAACCGGGCGGACGCGGCG 1260
OY 1261 TATAGCAAGACTTCCGCGCTCCGCTCTGGAAGAGACTGGTGAAGACTTTGAGAGC 1320
DB 1261 TATAGCAAGACTTCCGCGCTCCGCTCTGGAAGAGACTGGTGAAGAGACTTTGAGAGC 1320
OY 1321 ATATGAAAGCGCATCGAGAGTCTTCAAGCTCGAACCAGGCGCAGAGAGCGCGTT 1380
DB 1321 ATATGAAAGCGCATCGAGAGTCTTCAAGCTCGAACCAGGCGCAGAGAGCGCGTT 1380
OY 1381 GGAATTAAGGGGCAATGACTACGGGGGTTTATGACCAACCTGGGCTTGATCGAGCGAC 1440
DB 1381 GGAATTAAGGGGCAATGACTACGGGGGTTTATGACCAACCTGGGCTTGATCGAGCGAC 1440
OY 1441 CTCTTCAAGGCAAGATTAAGCGAAGAGCGCATAGCTAGCTACACAGCTACGCGCTTC 1500
DB 1441 CTCTTCAAGGCAAGATTAAGCGAAGAGCGCATAGCTAGCTACACAGCTACGCGCTTC 1500
OY 1501 TCGGACATATAGGGCTCTGTGACGACGTCGAGCTATCGGGCAATCCGTTAGGAACGAG 1560
DB 1501 TCGGACATATAGGGCTCTGTGACGACGTCGAGCTATCGGGCAATCCGTTAGGAACGAG 1560
OY 1561 AACCTCAGGAAGCTCAGCCGCTGTCTTACGCTCAGAGCTGAAGGGCGCATATCTCTA 1620
DB 1561 AACCTCAGGAAGCTCAGCCGCTGTCTTACGCTCAGAGCTGAAGGGCGCATATCTCTA 1620
OY 1621 ATCCACTCGCTTGAAGACTACCGCTGTCCGCTGACAGAGCCTTATGTCTTCAACGTT 1680
DB 1621 ATCCACTCGCTTGAAGACTACCGCTGTCCGCTGACAGAGCCTTATGTCTTCAACGTT 1680
OY 1681 CTCAGAGCATATGGGCAAGGAAGCTTACATATGAGATATTCAGCGCGGCGCCACGGCCAC 1740


```

DB 1681 CTCGAAGACATGGGCAAGGAGGCTCATATGCGATTTCAAGCGCGCGCCGCGCCAC 1740
QY 1741 AAGCTCCGGGGAAGCCGAGGACAGCGCGCTACAGGCTCTTCATAGAGTTCTTC 1800
DB 1741 AAGCTCCGGGGAAGCCGAGGACAGCGCGCTACAGGCTCTTCATAGAGTTCTTC 1800
QY 1801 GAGCGCACTCAAGAGATGAGGAGGCGCTTGTAGAGTACTCAAGGGAAT 1860
DB 1801 GAGCGCACTCAAGAGATGAGGAGGCGCTTGTAGAGTACTCAAGGGAAT 1860
QY 1861 GGGAACTGA 1869
DB 1861 GGGAACTGA 1869

RESULT 2
AAH41226/c
ID AAH41226 standard; DNA; 349980 BP.
XX
XX AAH41226;
AC
XX 29-OCT-2001 (first entry)
DE Pyrococcus abyssi genomic fragment #5.
XX
XX Hyperthermophilic archaeon; hyperthermophilic protein; ds.
OS Pyrococcus abyssi.

Key Location/Qualifiers
FH 1..49980
FT misc-feature
FT /tag= a
FT /note= "This sequence overlaps with the 3' end of
FT AAH41225"
FT 300001..349980
FT /tag= b
FT /note= "This sequence overlaps with the 5' end of
FT AAH41227"
XX
XX FR2792651-A1.
XX
XX 27-OCT-2000.
XX
XX 21-APR-1999; 99FR-0005034.
XX
XX 21-APR-1999; 99FR-0005034.
XX
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX (IFREMER INST FR RECH EXPL MER.
XX
XX Forterre P, Thiery JC, Prieur D, Dietrich J, Lecompte O;
XX Querellou J, Weissenbach J, Saurin W, Hellig R;
XX WPI; 2001-126236/14.
XX
XX New nucleotide sequences isolated from Pyrococcus abyssi encode
XX proteins useful in industry -
XX
XX Claim 1; Page 511-606; 1657pp; French.
XX
XX The present invention relates to the genomic sequence of Pyrococcus
XX abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a
XX hyperthermophilic archaeon, which is isolated from deep sea hydrothermal
XX vents. The present sequence is a fragment of the genomic sequence of P.
XX abyssi. The 5' end of this sequence overlaps with the 3' end of AAH41225
XX and the 3' end of this sequence overlaps with the 5' end of AAH41227. The
XX proteins of the present invention have various potential industrial uses,
XX since the proteins are stable at very high temperatures, some up to 110
XX degrees centigrade.
XX
XX Note: This patent is in the same patent family as WO200065062, which
XX contains additional sequences as shown in AAB99132-AAB99143,
XX AAH75903-AAH75920 and AAG66436.

```

```

XX SQ Sequence 349980 BP; 98084 A; 80447 C; 77665 G; 93784 T; 0 other;
Query Match 57.7%; Score 1078.2; DB 22; Length 349980;
Best Local Similarity 74.9%; Pred. No. 5.9e-260;
Matches 1405; Conservative 0; Mismatches 458; Indels 12; Gaps 4;

QY 1 ATGACCGGCGATGATGAGAACACAGACCTTTTAAATGCGCTACCTGCGGACCGG 60
DB 206185 ATGACCTCATATGAAATGAAATGAAAGAGCTTTACAAATTTGCTTACCTAGCCACCCC 206126
QY 61 AGGATACGGGGAACTTAATCGGTACACCTGACGAAAGGCAATGAAAGACAACAAG 120
DB 206125 AGGATACGGGGAAAGCAATATACGCTTCTAACCAAGGCACTTACAAACAACAAG 206066
QY 121 TAGGAGAGCGGTTGTGTGAAGCCTTGAAGCGGCTCAAGGCGCTTCAATCGAAGAC 180
DB 206065 TAGGAGAGCGGATGTTCTCGAAGACCTTCGAGAGATGGAGCAAGAGTTCAATGAAAGAC 206006
QY 181 GCGTCATATGCCGAGATTTGCGCAGAGCGGACAGAAAGCTTCGCTTCACTGCTTTAAGAG 240
DB 206005 GCTTCATATGCTAGATTTTCTCAGATGGAAGAAAGATAGCTTCTATGAGTTCAACGAG 205946
QY 241 GAGAAAGAGAGAACCGAGATATGGTGGCCATATCCAGACCTGAGCGCAAGAAAGTC 300
DB 205945 GAGAAAGAAAGTCTCAGATATAGGGTGGCGGACATGAAAGACCTTAAGCGCAAGAAAGTT 205886
QY 301 CTCTCAACTAAAAAGTCGCGCTCGATGAGTGAAGCAAGATGAAAGGAGACTCTACTT 360
DB 205885 CTGAGGCCCAAGAAATAGTACCTCGAATGGAATGAGACTCAAGAGAGTTGTTAAT 205826
QY 361 GTGCGCTTCAAGAGAGGAGAGAGTGAAGACTTCTTGTGACGACGACCTCCGCTGCG 420
DB 205825 ATAGGCTTCAAGAGAGAGAGAGATGAATTTCAATATTCAGAGATGAGTTCACACTTGG 205766
QY 421 TTGCAATATATGAGATTTCTTGTATGAGAGAGAGAGAGACTTCTGCGTCTTGACACTGAG 480
DB 205765 TTGCAATATATGAGATTTCTTGTATGAGAGAGAGAGAGACTTCTGCGTCTTGACACTGAG 205706
QY 481 GCCGAGAGATATATGAGACAGTTGAGAGAACCGGAGCTTTGAGAGGCGCTCGGACAGGC 540
DB 205705 GCCGAGAGAGATATATGAGACAGTTGAGAGAACCGGAGCTTTGAGAGGCGCTCGGACAGGC 205646
QY 541 GATCGGATAGTTGTGAACGTCGCCGACCGAGAGGAGGAGCAAGCTGCTTCAAGTTTC 600
DB 205645 GATTCGATAGTTGTGAACGTCGCCGACCGAGAGGAGGAGCAAGCTGCTTCAAGTTTC 205589
QY 601 TACGACATATGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 205588 TGAACATCTACCTCTGGAAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 205529
QY 661 GAGCGGTTGACTCCGAGAGAGAGAGATATCTCCGAGGAGCAAGAAAGAGAGAGAGAGAG 720
DB 205528 CAGCGTATATGACTCCGAGAGAGAGAGATATCTCCGAGGAGCAAGAAAGAGAGAGAGAGAG 205469
QY 721 ATGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 777
DB 205468 ATGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 205409
QY 778 GGCCCGGCTGACGCTGTGGAGAGCAAGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 837
DB 205408 ATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 205349
QY 838 GATCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 897
DB 205348 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 205289
QY 898 GACACATGATTTACGGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 957
DB 205288 AAGCATGAGATATATGAGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 205232
QY 958 GCCACGAGAGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1017

```


OY 1141 ATAGCTTCGTCACGGGGGCGGAGGCGATGTACGACACCGCTTCTACGATG 1200
 DB 58645 GTGCTGGAAATTCACGGTGGGCGCAAAACAGCCTACGCTTATGACAGAGTTT 58586
 OY 1201 CAGCTGATGCGAGCAAGGCTACTACTGCTGCTGTAACCCCGGGGCGAGCAGGC 1260
 DB 58585 CAGGTTTATGTTGCCAAGAGGCTTCTGTTATCTTTTCAATTCACAGGGAGATGATG 58586
 OY 1261 TATGACGAGACTTCGCGCTCCGCTCTGAGAGAGACTGCTTGGAGACTTTGAGGAC 1320
 DB 58525 TATGCTGAGAGTTGCGG---GACATAGGGGTCTACAGCGTGAGCGGATATATCAAG 58469
 OY 1321 ATATGAAAGGCGATAGGAGTCTTTCAGACTCGAACCCGCGAGCGAGAGCGCTT 1380
 DB 58468 CTAATGAGGATAGTACGAGGCTTTAAAGCTTGTATGATTAAGAGAGAAAGCGCTT 58409
 OY 1381 GGAATTAACGGGCAATAGCTACGGGCTTCTATGACCACTAGGCGCTTACTCAGAGCGAC 1440
 DB 58408 GCGCTTACCGGAGATCATATGAGAGGCTTATGACAAATTTGGATCGTTGGGCGACAAAC 58349
 OY 1441 CTCTTCAAGGAGAAATPAGGAGAAACGGCATTAAGCTACTGCTCACCAGCTACGCCCTTC 1500
 DB 58348 AGATTCAAAAGCTGCGGTAACACACGCTCAATATCCAACTGATCAGCTCTTTGGAACG 58289
 OY 1501 TCGGACATAGGGCTCTGTGTAAGCTGAGGTCATCGGGCCCAATCCGTTAGAAC--- 1557
 DB 58288 ACCGATATCGGCTATTTACTTCTGCTCCAGATAGATCGGTGCGATCTTGGAGTAACTTA 58229
 OY 1558 GAGACTTCAAGAAAGCTCAGCCGCTGTTACGCTCGAAGAGGAGGCGCGATACTC 1617
 DB 58228 GAAAGCTACTGCGGAGAGAGCCATTTGAATAGCTCCAAAGTTGAGACCTCCGCTCTG 58169
 OY 1618 CTATCACTGCTTGGAGACTACCGCTGCTCCGCTCGAACGAGCTTATTTTACAAAC 1677
 DB 58168 ATATTTACTCTCCACGAGACTACCGCTGCTTCCGAGGCTTTCGCAATTTTACATA 58109
 OY 1678 GTGCTCAAGACATAGGCGCAAGAGCTTACATAGCATATTCAGACGGGCGGCGACGCG 1737
 DB 58108 GCGTTGAAGTATTTGGCGCAAAACCGTGAAGTGGCGATATTTCCAGGTGAGAAATCATGAC 58049
 OY 1738 CACAGCTCCGCGGAGAGCCGAGGACAGCGGCGGAGGCTTCTTCAATAGAGTTC 1797
 DB 58048 TTGAGTGAAGTGAAGGCAAAAGCATAGAGGTTAAGAGGCTCGAGCTCATATGTGGGG 57989
 OY 1798 TTGAGGCGCAAGCTCAAG 1815
 DB 57988 TTGAGAGGTGGCTGAAG 57971
 RESULT 4
 ID AAX07301 standard; DNA: 1971 BP.
 XX AAX07301:
 AC 21-MAY-1999 (first entry)
 DT Bacillus subtilis serine protease SP1 (Yuxl) DNA.
 DE Serine protease SP1: Yuxl; detergent: ss.
 KW Bacillus subtilis.
 OS Bacillus subtilis.
 PN WO9003984-A2.
 XX 28-JAN-1999.
 PD 14-JUL-1998; 98MO-US14647.
 PF 15-JUL-1997; 97EP-0305232.
 PR (GENV) GENENCOR INT BV.
 XX (GENV) GENENCOR INT INC.
 PA

XX Estell DA;
 PI WPI: 1999-132231/1.
 DR P-PSDB: AAM97789.
 XX
 PT Use of serine protease genes from Gram-positive microorganisms - for
 PT modification of host cells for the production of heterologous
 PT proteins or for producing proteins for use in cleaning compositions
 PS Disclosure: Fig 1A-C; 37pp; English.

CC This DNA sequence encodes serine protease SP1 (Yuxl) of Bacillus
 CC subtilis. Novel serine proteases SP1, SP2, SP3, SP4 and SP5 (see
 CC AAM97789-93) were identified via a FASTA search of Bacillus subtilis
 CC genomic nucleic acid sequences. SP1 was identified by its
 CC structural homology to the S9 type serine protease dbp2 of yeast.
 CC SP2, SP3, SP4 and SP5 were identified by their structural and
 CC overall amino acid homology to SP1. Host cells in which the
 CC naturally occurring gene encoding one or more of SP1, SP2, SP3,
 CC SP4 or SP5 is mutated such that the proteolytic activity is
 CC diminished or deleted altogether, can be used for the production of
 CC heterologous proteins, e.g. a hormone, enzyme, growth factor,
 CC cytokine, protease, carboxylase, lipase, racemase, epimerase,
 CC tautomerase, mutase, transferase, kinase or phosphatase (claimed).
 CC SP1, SP2, SP3, SP4 and SP5 can also be produced on a large scale
 CC in a microbial host expression system for use in cleaning
 CC formulations such as detergents, bar or liquid soap, dish-care
 CC compositions and contact lens cleaning solutions, or for peptide
 CC hydrolysis, waste treatment, textile applications, as
 CC fusion cleavage enzymes in protein production, and as animal feed
 CC additives.

Sequence 1971 BP: 558 A; 431 C; 524 G; 458 T; 0 other;

Query Match 8.0%; Score 149.8; DB: 20; Length 1971;
 Best Local Similarity 52.3%; Pred. No. 1.4e-27;
 Matches 356; Conservative 0; Mismatches 322; Indels 3; Gaps 1;

OY 1135 CCGGTGAAAGCTCTGTCACAGGGGCGGAGGCGATGACGAGCGGCTGCTGCTAC 1194
 DB 1291 CCACTTATTTCTTACATATACAGCGGCTCCCATATGATGACGACATATATTTTCA 1350
 OY 1195 GAGATGCACTGATGCGAGCAAGGCTACTACTGCTGCTGTAACCCGCGGCGAGC 1254
 DB 1351 GAGTTTCAAGTGTCTGGCGGGAAGGATGACGCGTCTTATATCATCCGAGAGGAA 1410
 OY 1255 GAGGCTATAGCGAAGACTGCGGCTCCGCTCTGAGAGAGACTGCTGAGAGCTTT 1314
 DB 1411 CAGGCTACGGGCGAGAAATTTGATGCGGTACAGAGATTAATGAGGGAAGGATTA 1470
 OY 1315 GAGGACATTAATGAGAGGATATGAGAGTCTTCAAGCTCGAACCGGAGCGAGAGGAT 1374
 DB 1471 GAGATGATGATGACGCTGTGAGAGAGGCTATCAAGAGATCCGATATGATCCTAAG 1530
 OY 1375 CCGCTTGAATTAACGGGCAATAGCTACGGCGCTTCAATGACCAATGCGGCTGACTGAC 1434
 DB 1531 CCGCTGCTGTCACGCGGAGAGGCTACGAGGTTTATGACCAACTGAGTCTGCGGCGAG 1590
 OY 1435 AGGACCTCTTCAAGGAGGAGATTAAGGAGAGAGGCGCATTAAGTACTAGGCTCACAGCTAC 1494
 DB 1591 ACGAACCGCTTTAAGTATGCTCCGCTTACAGAGCTCGATATCAAAATTTGATCAGCTTAC 1650
 OY 1495 GCGCTTCCGACATTAAGGCTCTGTAACGAGCTGAGAGTCAATCGGCGCAATTCGTTAGAG 1554
 DB 1651 GCGCTCAGTATTAAGGCTATTTCTTACAGAGCTGAGAGCTTGAGATGATGTTGAG 1710
 OY 1555 AAC---GAGAACTTCAAGAAAGCTCAGCCCGCTGTTCTACGCTCAAGAACTGAAGCGCGG 1611
 DB 1711 GACACAGAAAAGCTCTGGGACCGCTCTTTAAATATGACCAACCAAGAGAGACACG 1770
 OY 1612 ATACTCTATTCACATCGCTTGAAGACTACCGCTGCTCGAGCAAGAGCTTATGTTTC 1671

Key	Location/Qualifiers
FF	1.1896
FF	/*tag= a
FF	/product= "Pyrococcus horikoshi acylpeptide hydrolase"
FF	/note= "CDS does not contain a stop codon"
PD	JP10210977-A.
XX	
XX	11-AUG-1998.
XX	
XX	31-JAN-1997; 97JP-0018381.
XX	
XX	31-JAN-1997; 97JP-0018381.
PA	(AGEN) AGENCY OF IND SCI & TECHNOLOGY.
DR	WP1; 1998-488369/42.
DR	P-PSDB; AAM70508.
XX	
XX	A heat-resistant acylpeptide hydrolase and a gene coding it - useful
XX	for hydrolysing the C terminl of proteins at high temperature
XX	
XX	Example 5; Pages 6-7; 8pp; Japanese.
XX	
XX	The present sequence represents a Pyrococcus horikoshi acylpeptide
XX	hydrolase encoding DNA sequence. The enzyme has the following
XX	properties: (a) it hydrolyses an acylpeptide; (b) an optimum temperature
XX	of 90-95 degrees Centigrade; (c) an optimum pH of 5.0-6.0; (d) no loss
XX	of activity when heated to 95 degrees Centigrade for 3 hours at pH 7.5;
XX	and (e) a molecular weight of 60 kDa. The enzyme is claimed useful
XX	for hydrolysing the amino end of an acylated protein and a peptide at
XX	high temperatures.
XX	
XX	Sequence 1896 BP; 614 A; 358 C; 491 G; 433 T; 0 other;
XX	
XX	Query Match 6.4%; Score 120.4; DB 19; Length 1896;
XX	Best Local Similarity 50.3%; Pred. No. 3.3e-20;
XX	Matches 351; Conservative 0; Mismatches 341; Indels 6; Gaps 2;
QY	1120 AAGAGGAGGAAGGCCCGGTGTATGCTTCCTCCACGGCGGGCCGAAGGCGATGACGGA 1179
DB	
QY	1198 AAGAGGAAGAGTATTCAGGATATCTAGAGATCCACGGTGGTCTTAAACCGCTTACGGT 1257
QY	1180 CACCGCTTCGCTACGAGATGCACGCTGATGGCGAGCAAGGCGTACTACTGCTGCTG 1239
DB	
QY	1258 TACGCTTTTATGACGAGAGTTCACGCTTTTAACTCTAAAGGCGTTCGTGATATTCGA 1317
QY	1240 AACCGGCGGCGACGAGCGCTATATAGGAGAACTTCGGCTCCGCGTCTGGAAGAGACT 1299
DB	
QY	1318 AATCTCTAGAGGAGCGATGGCTACGAGAGGAGATTCGGCGATA--TAAAGGACACTAT 1374
QY	1300 GCGTTGAGGACTTTGAGGACATATATATTAACGGCATCGAGAGTCTTCAACCTGGAACG 1359
DB	
QY	1375 GGGGAGAGGGATTACAGAGATTTATTTAGAGTATTCATGAAGCATTTAAGAGATTGGAC 1434
QY	1360 CAGGCCACAGAGGAGCGCGTTGGATTAACGGGCAATTAAGCTACGGCGGCTTCATGACCAAC 1419
DB	
QY	1435 TTCAATATATGGGGAAGGCTAGGAGTAAACGGGGGGTTCCTATGGTGGCTTCATGACGAAC 1494
QY	1420 TGGGCGCTTGACTGAGAGCACTCTCTTAAGGCGAAGATTAACCGGAACGCGATTAAGCTAC 1479
DB	
QY	1495 TGGATTAAGCTCGGACATTCACCAAGGTTAAAGCCCGCTGTAAACCCAGAGATCAATTTCAAT 1554
QY	1480 TGGCTCAACGACTACGCTTCTCTCGAGATATAGGGGCTGTGATCGAGAGTTCGAGAGTATCGGG 1539
DB	
QY	1555 TGGATTAAGCTTCTCTCGGAGACAGCGATATTAAGTATTACTTTGCTCCAGATCAATTTAGGA 1614
QY	1540 CCAATATCCGTTAGAGAAC--GAGAACTTCAGAGAGCTCAGCCCGCTGTTCTACGCTCAG 1596
DB	
QY	1615 AAGATATCCGAGCAACTTGGAGGTTATTTGGAAAAAGGCCCATTTAAGATGACGCTCC 1674
QY	1597 AAGGTGAAGCGCGGATATCTATATTCACACTCGCTTAAAGGACTACCGCGTCCGCTGCAC 1656

[illegible]

CC (1) are useful for modulating the polynucleotide content or composition
 CC of an organism. (1) is useful for identifying, isolating or synthesizing
 CC DNA molecules such as promoter, DNA binding elements, open reading frames
 CC or full-length genes, that then can be used as expressible DNA in
 CC transgenic organisms. (1) may be used to detect lactic acid bacteria,
 CC preferably *L. rhamnosus* in a sample material. (1) is also useful for
 CC genome mapping, physical mapping, and in positional cloning of genes of
 CC more or less related microbes, and to design probes and primers. (1) is
 CC also useful for transforming microbes for use in a therapeutic
 CC composition that is effective for treating or preventing a
 CC gastrointestinal condition or disorder caused by the presence of
 CC pathogenic microbes in the gastrointestinal tract or by the absence of
 CC normal intestinal microbes in the intestinal tract. Proteins are used to
 CC raise antibodies, to isolate corresponding interacting proteins, as
 CC nutritional additives and as additives in dairy processing and
 CC fermentation and processing. (1) and encoded proteins are used for the
 CC selection and production of more effective probiotic bacteria, as
 CC bioactive (health promoting) ingredients and health supplements, for
 CC immune function enhancement; for reduction of blood lipids such as
 CC cholesterol; for production of bioactive material from genetically
 CC modified bacteria as adjuvants; for wound healing; in vaccine
 CC development, in selection and production of genetically modified rumen
 CC microorganisms for improved animal nutrition and productivity, better
 CC flavour and improved milk composition.

XX Sequence 1032 BP; 252 A; 220 C; 281 G; 279 T; 0 other;

Query Match 4.18; Score 76.6; DB 24; Length 1032;
 Best Local Similarity 48.7%; Pred. No. 2.7e-09;
 Matches 208; Conservative 0; Mismatches 219; Indels 0; Gaps 0;

DB 1085 TCGAGATAGACGGCTGTACTTCCAGGCGCGAGGTTAAAGAGGAAAGCCCGGTGTAG 1144
 263 TTGAGATAGGCGCTGTATTTCCACCGCAACAGCGCTATCATCGCAATCGGCAATTT 322
 QY 1145 TCCTTCGTCACGGCGCGGCGCAAGGCAATGACGACACCGCTTCGTACAGATGACG 1204
 323 TGTATGTCATGCGCGCGCGCGCGAGTCGATATGCTTATTCATGAATATGAGT 382
 DB 1205 TGTATGCGAGCAAGGCTACTACTGCTGCTTGTGAACCGCGCGCGAGGAGGCTATA 1264
 383 ATCTGCGACGAAAGGCTATGAGCGGTATTTGTGAATTCGCGTGGAGGCTTATGATACC 442
 QY 1265 GCGAAGACTTCGCGCTCCGCTCCTGTGAGAGAGACTGCGCTTGAAGGACTTGAAGCACTAA 1324
 443 GCGAGGACTTTACGGCGCGCTGTATTAACATTAACCGGAGCGGATTAAGATTTGCT 502
 DB 1325 TGAAGGCAACGAGAGATTTCTCAAGCTCGAACCGGACGCGAGGAGCGGCTTGAA 1384
 503 TGGCTTCGGGTGAAGGCGCTTAAGCTCGATACAAATTAATCCGCAACGCTTATTTG 562
 QY 1385 TAACGGGCAATACCTACGCGGCTTCAATGACCAACAGGCGCTTACTCAAGAGACTCT 1444
 563 TCACCTGCGGCTTCTTATAGCGGCTTATGACTAATGATGATGACCAATACGCAATGCTT 622
 DB 1445 TCAAGGCAAGATTAAGGCAACGCGATTAAGCTACTGCTACACAGCTACGCTTCTCG 1504
 623 TTAAGAGAGGGAATACCAACGCTTCATTTCCAAATGGCGTAGATATATGATACGACGT 682
 QY 1505 ACATPAG 1511
 DB 683 ACATCGG 689

RESULT 8
 ABK77575

ID ABK77575 standard; DNA; 477 BP.

XX ABK77575;

DF 13-AUG-2002 (first entry)

XX *Bacillus clausii* genomic sequence tag (GST) #418.

XX Differential gene expression; genomic sequenced tag; GST;
 KW altered culture condition; environmental stress;
 KM physiological provocation; ds.
 XX *Bacillus clausii*.
 OS WO200229113-A2.
 PN 11-APR-2002.
 PD 05-OCT-2001; 2001MO-US31437.
 PF 06-OCT-2000; 2000US-0680598.
 PR 27-MAR-2001; 2001US-279526P.
 XX (NOVO) NOVOZYMES BIOFTECH INC.
 PA (NOVO) NOVOZYMES AS.
 XX Berka R, Clausen IG;
 PT WPI; 2002-416684/44.
 PT Monitoring differential expression of several genes in first *Bacillus*
 PT cell relative to expression of same genes in one or more second
 PT *Bacillus* cells, by using substrate containing *Bacillus* genomic
 PT sequenced tag array -
 PS Claim 11; SEQ ID NO 4866; 200pp; English.

CC The invention describes a method of monitoring differential expression of
 CC genes in a first *Bacillus* cell relative to expression of the genes in
 CC other *Bacillus* cells, comprising hybridising labelled nucleic acid probes
 CC isolated from *Bacillus* cells to a substrate containing array of *Bacillus*
 CC genomic sequenced tags (GST), examining the array, and determining
 CC relative gene expression by an observed hybridisation reporter signal of
 CC a spot in the array. The method is useful for measuring the expression of
 CC genes in a first *Bacillus* cell relative to expression of the same genes
 CC in one or more second *Bacillus* cells. The method is useful for monitoring
 CC global expression of several genes from a *Bacillus* cell, discovering new
 CC genes, identifying possible functions of unknown open reading frames and
 CC monitoring gene copy number variation and stability. Monitoring changes
 CC in expression of genes may be used to provide a representation of the way
 CC in which *Bacillus* cells adapt to changes in culture conditions.
 CC environmental stress or other physiological provocation. Extensions
 CC follow-up characterisation is unnecessary, when one spot on an array
 CC equals one gene or one open reading frame, since sequence information is
 CC available. This sequence represents a genomic sequence tag (GST) used in
 CC the method of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

XX Sequence 477 BP; 138 A; 89 C; 118 G; 132 T; 0 other;

Query Match 3.58; Score 66.2; DB 24; Length 477;
 Best Local Similarity 49.9%; Pred. No. 8.6e-07;
 Matches 167; Conservative 0; Mismatches 168; Indels 0; Gaps 0;

QY 1366 GACAGGAGGCGGCTTGAATTAACGGGCTAAGCTACGCGCTTATGACCAATGCGCC 1425
 30 GATCAGGAGGCGGCTGCTGCTGAGGGAAGTTAGCGTGTATTAAGCAAAATTTGGCT 89
 DB 1426 TTGACTAGAGCAGCACTCTTCAAGCAGGAATTAAGCAGAGAGCGCAATCTAGCTC 1485
 90 GTTGGCCATAGCAATCGGTTCAAGCGCGCTTACGCAACGCTGCAATCAATTTGGATT 149
 QY 1486 ACCAGTACGCGCTTCTGAGCATAGAGGCTGTAGCAGCTCAGAGTCAATGCGCAAT 1545
 150 AGTTTATAGCGGCTAAGCAATATTTGCTATTTTCAAGAGTGGCAATTAAGCAGAC 209
 QY 1546 CCGTTAGAGAGAGAACTTCAGGAGACTACGCCGCTGTTCAGCTCAGAAAGTGAAG 1605

Db	210	TTGCATGATATCGAAGAAACACTGTGGGCCCATTCGCGACATCGAATACGTTGAAACATCGGGA	269
Oy	1606	GGCGCGAATACCTCTAATCCACTCGCTTGAGGACTACCGCTGTCCCTGGACCAAGACCTT	1665
Db	270	ACGCCATTGCTTATTCCTACAGCGTGAAAAAGATTTCCTCCCTGTCCATTCGACACACAGAG	329
Oy	1666	ATGTTCTACAAACGTGCTCAAGACATGGGCAAGGA	1700
Db	330	CAATGCTTATTCGCTTGAAAAAGCATGGAAAGA	364
RESULT 9			
AAFe0975			
AAFe0975	standard; DNA; 1836 BP.		
AAFe0975;			
16-MAY-2001	(first entry)		
DE	P. putida KT2440-associated DNA ORF00863.		
DE			
XX	Transgenic plant; detection; probe; amplification; vaccine carrier;		
KW	microbial production strain; biological remediation; ds.		
XX			
OS	Pseudomonas putida.		
XX			
PN	DE19935088-A1.		
XX			
PD	01-FEB-2001.		
XX			
PE	27-JUL-1999; 99DE-1035088.		
PR	27-JUL-1999; 99DE-1035088.		
PA	(TIGR-) TIGR INSTR GENOMIC RES.		
PA	(ODIA-) ODIA GEN GEN.		
PA	(GHRP) GRS BIOTECHNOLOGISCHE FORSCHUNG MBH.		
PA	(DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.		
XX	(MED1-) MEDIZINISCHE HOCHSCHULE HANNOVER.		
DR	WPI; 2001-192469/20.		
XX			
PT	New DNA sequences specific for Pseudomonas putida KT2440, useful as		
PT	sate genetic engineering host, allow detection in presence of other		
XX	related bacteria -		
PS	Claim 1a; Page 21-22; 158pp; German.		
CC	This invention describes novel DNA sequences (I) for specific detection		
CC	of Pseudomonas putida KT2440. The invention also describes (1)		
CC	recombinant expression vector containing (1); (2) prokaryotic or		
CC	eukaryotic cells transformed or transfected with (1) or the vector of		
CC	(1); (3) production of expression products by culturing cells of (2);		
CC	(4) expression products, or their fragments, of (1) and synthetic		
CC	proteins or peptides with the same sequences (A); (5) poly- or		
CC	mono-clonal antibodies (Ab) that react specifically with (A); (6)		
CC	hybridoma cells that produce the monoclonal Ab of (5); (7) transgenic		
CC	plants that contain transformed or transfected cells of (2); (8)		
CC	detecting KT2440 using a labeled (1) or Ab as probe; and (9) DNA chips		
CC	carrying one or more (1), (1), and their fragments, are used as probes		
CC	to detect and isolate full-length cDNAs and/or to amplify such cDNAs by		
CC	polymerase chain reaction, and for production of transgenic plants, (1),		
CC	or antibodies that recognize their expression products, are used for		
CC	detecting the presence of KT2440, particularly in presence of other,		
CC	even closely related, bacteria. KT2440 is one of the bacteria classified		
CC	as safe, by the National Institutes of Health, for genetic engineering		
CC	work, e.g. as microbial production strains, for biological remediation		
CC	and as vaccine carriers. (1) are exclusive to KT2440 with no significant		
CC	homology with sequences in other bacteria (specifically the closely		
CC	related pathogen P. aeruginosa). Compared with other 'safe' bacteria, it		
CC	has greater catalytic activity and better survival in, and adaptation to,		
CC	the rhizosphere and soil.		

OS Pseudomonas aeruginosa.
 XX
 PN W0200170955-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-US09180.
 XX
 PR 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-259308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Haselbeck R, Ohlsen K, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX
 DR WPI; 2001-611495/70.
 DR P-PDB; AA036516.
 XX
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 PS Claim 27; Seq ID No 8012; 511pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence encodes an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 783 BP; 183 A; 268 C; 224 G; 108 T; 0 other;
 XX
 Query Match 2.9%; Score 55; DB 23; Length 783;
 Best Local Similarity 45.1%; Pred. No. 0.00066;
 Matches 205; Conservative 0; Mismatches 250; Indels 0; Gaps 0;
 XX
 QY 882 GCGGTGTTGTTACCGGACACATGATTACGGGCTTGAGCGATGCAAGCAATC 941
 DB 12 GCTTGGCTGTTTCGCGCTGCGGGGCTCGCGGCTCAGCGCGCCAGCCGCCGAGTC 71
 QY 942 GCTTCATCATATACCGCCGACGAGTAGCGGAGCTTACTTACGAGCGCGCAAGCTGAA 1001
 DB 72 CTTCCACCGTGGCGGACCCCGGTGGCGGACGAGAGTCTTACGCTGTCAGAGCGCT 131
 QY 1002 ACAGGTACCGGATTAACAGGGGCGGATATTCAGAGAGCTCAAGCTTCGAGCGGAGCA 1061
 DB 132 GCTGGCCAGGAAGGCGGTGACCTGAAGATCAAGAGTTCACGATCTGACGCCGAA 191
 QY 1062 CTTCCGCTTACAGCAAGACCTCGAGATAGACGGCTGTACTTACAGCGCGGAGGTTAA 1121
 DB 192 CGTGACAGGCTCGGAAAGGCGCTGAGACGCACTTCTTCACACACAGCCGTAACCTGCA 251
 QY 1122 AAGAGGAAGGCCCCGGTGTACTCTTCGTCACGGCGGCGCAAGAGGCAATGACGACA 1181

DB 252 TGAGTTCAACAAGGCCAAGGGCAACGACCTGTGTGGCGTACCGGGTACACATGAGCC 311
 QY 1182 CCGCTTCGCTCTACAGATGACGTATGCGGACCAAGGCTACTACTCTTCTGTAA 1241
 DB 312 GCTGGGCGGCTTACTGTGAGCAAGATGACGAACTCTCCGGCGCTTACCGCT 371
 QY 1242 CCGCGCGGAGGAGGCGCTATAGCGAAGCTTCGCGCTCCGCTCTGAGAGGAGCTG 1301
 DB 372 GGTGATTTCCCAAGGACGACCAAGCGGCGCGGCTGTCTCTGTGACAAAGCGCG 431
 QY 1302 CTTGAGGAGCTTTGAGGACATTAATGAACGGCATCG 1336
 DB 432 GGTGATCAAGCTCAAGGACCAACAAGAGCATCACCG 466
 XX
 RESULT 11
 AAF13721
 ID AAF13721 standard; cDNA; 659 BP.
 XX
 AC AAF13721;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Aspergillus oryzae EST SEQ ID NO:6244.
 XX
 KW Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS Aspergillus oryzae.
 XX
 PN W0200056762-A2.
 XX
 PD 28-SEP-2000.
 XX
 PF 22-MAR-2000; 2000WO-US07781.
 XX
 PR 22-MAR-1999; 99US-0273623.
 XX
 PA (NOVO) NOVO NORDISK INC.
 PA (NOVO) NOVO NORDISK AS.
 PI Berka RM, Rey MW, Snustner JR, Kauppinen S, Clausen IG, Olsen PB;
 XX
 DR WPI; 2000-594572/56.
 XX
 PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 XX
 PS Claim 88; Page 2569; 3161pp; English.
 XX
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate

ID AAA12989 standard; DNA; 1062 BP.
 AC AAA12989;
 DT 18-JUL-2000 (first entry)
 XX
 DE DNA encoding Cellulomonas fimi xylanase.
 XX
 KW xylanase; endo-1,4-beta-xylanase; xylan hydrolysis;
 KM xyl-o-oligosaccharide production; ds.
 XX
 OS Cellulomonas fimi.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1062
 FT /tag- a
 FT /product- "Cellulomonas fimi xylanase"
 XX /note- "No stop codon given in the specification"
 XX
 PN JP3030331-B1.
 XX
 PD 10-APR-2000.
 XX
 PF 17-MAR-1999; 99JP-0071715.
 XX
 PR 17-MAR-1999; 99JP-0071715.
 XX
 PA (NORC) NORIN-SUISANSO SHOKUHN SOGO KENKUSHOCHO.
 PA (SEIB-) SEIBUTSUKET TOKUTETI SANGYO GIUTSU KENKYU SUISHIN KIKO.
 DR WPI; 2000-306509/27.
 DR P-PSDB; AAY81496.
 XX
 PT Modified xylanase gene for use in foodstuff industry, has base sequence
 of 1065 nucleotides -
 XX
 PS Example 1; Page 9-10; 11pp; Japanese.
 XX
 CC The invention relates a novel chimeric xylanase (AAY81494) and to DNA
 encoding it (AAA12985). The chimeric xylanase comprises residues 1-244
 of Streptomyces olivaceoviridis xylanase and residues 244-354 of
 Cellulomonas fimi xylanase. Xylanase (also known as endo-1,4-beta-
 CC xylanase) hydrolyses beta-1,4-D-xylan, a component of the hemicellulose
 in plant cell walls, into xyl-o-oligosaccharides and xylose. Xylanase is
 used in a range of industrial processes. It is used to produce xyl-o-
 CC oligosaccharides from xylan from broad-leaved trees, and is used in
 wood pulp bleaching to reduce the amount of chlorine required for this
 CC process. Xyl-o-oligosaccharides can be used as ingredients in foodstuffs
 and as water-retaining material in cosmetics. The chimeric xylanase of
 CC the invention does not generate xylose monomers during the hydrolysis of
 CC xylan. It is therefore useful for efficient and reliable xyl-o-
 CC oligosaccharide production. The present sequence represents DNA
 encoding xylanase from Cellulomonas fimi.
 CC
 SQ Sequence 1062 BP; 181 A; 388 C; 358 G; 135 T; 0 other;
 XX
 XX
 Query Match 2.5%; Score 47.4; DB 21; Length 1062;
 Best Local Similarity 43.3%; Pred. No. 0.059;
 Matches 222; Conservative 0; Mismatches 291; Indels 0; Gaps 0;
 OY 829 CTCACATCCAGATCGGGGCAAGGTAAACCTCTGCTGTGGACGGGAAGCCGAGCGTGT 888
 DB 67 CTCGCCGCGCGCGCGGACGCTGCTGCGGCGCCACGCGTGTGCTGCCGCCGACAGCC 126
 OY 889 GTTACCGGCGACCACTGATTTACGGGCTTGAGCTGACGATGCAAGATGCTTCCTC 948
 DB 127 GCGACCAAGCTCTAAGAGAGCGCGGCGGCGGCGGCGGCTTCGCTTCGCTGAC 186
 OY 949 ATCATGACGCGCACGAGATAGGCGAGCTTACCTTACGACGCGGAGCTGAAGAGTGC 1008
 DB 187 CCCAACCAGCTCTCGGAGGCGGACGTACAGGCGATGCCGACACCGAGTTCAACCTC 246
 OY 1009 ACCGAATACAAAGCGGCGGATATTTCAGGAAGCTCAAGACCTTTCGAGCGACCTCCG 1068

DB 247 GTCCGCGAGAACCGGATGAAGTGGAGGCCACCAGCCCTCCGACAGACTTCCTTC 306
 OY 1069 TTCAGAGCAAAAGACCTCGAGATAGACGGCTGGTACCTCAGCCGAGTTAAAGAGG 1128
 DB 307 GCGCGGCGGACCGCGCTCGAGCTACGCGCGGCGGACACCGGCAAGAGAGCTTACGCGCAC 366
 OY 1129 AAGCCCCGGTATATCTTCCTCCAGCGCGGCGGCGGAGGAGATGTAGGAGACCGCTTC 1188
 DB 367 ACGCTGTATGAGCTGCGACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 426
 OY 1189 GTCTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1248
 DB 427 GAGAGCGCATGTGTCAACACAGTACAGGATGATGATGATGATGATGATGATGATGATGATGAT 486
 OY 1249 GGCAGCGGCGGCTATAGGAGACTTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 1308
 DB 487 TCGTGGGAGCTGCTCAACGAGCGGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 546
 OY 1309 GACTTGGAGACATATGAAGCGCATCGAGAG 1341
 DB 547 TTCACAGAGAGCTCGGCAAGCGCTACATCGAG 579
 RESULT 14
 ID ABA47739
 AC ABA47739 standard; DNA; 1785 BP.
 XX
 AC ABA47739;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Human breast cell single exon nucleic acid probe #6434.
 XX
 KW Human; microarray; single exon probe; gene expression; breast;
 KW disease; cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200157271-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-0500662.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DT WPI; 2001-496933/54.
 XX
 PT New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes -
 XX
 PS Claim 4; SEQ ID NO 6434; 327PP + sequence listing; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BT 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for

CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 1785 BP; 293 A; 622 C; 480 G; 390 T; 0 other;

Query Match 2.5%; Score 47; DB 22; Length 1785;

Best Local Similarity 49.4%; Pred. No. 0.089;

Matches 122; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 1062 CTTCGGCTTCAGAGCAAGACCTCGAGATAGAGCGCTGTACTCAGCGCGAGGTTAA 1121
DB 510 CATGCCCTTGAATTGGCCCTACCTGAGAGTCTCTGGCCAGACCTTCATCACTGCTGT 569
QY 1122 AGAGGAG 1181
DB 570 CGGCCACCTGTGTCTCTCAACGACGACGCTCCGCTGCTCTATGTCTACCTGCTCTA 629
QY 1182 CCGCTTCGTACGAGATGAGCTGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1241
DB 630 TCTCTTCTTCGAGTGCACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 689
QY 1242 CCGCGCGGAG 1301
DB 690 CTACCTGTGTGTCTCTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 749
QY 1302 CTTCGAG 1308
DB 750 CCTGGGG 756

RESULT 15

ABA65633 standard; DNA; 1785 BP.

XX ABA65633;

XX 01-FEB-2002 (first entry)

XX Human foetal liver single exon nucleic acid probe #13938.

XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

XX MO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001MO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human fetal liver -

XX Claim 4; SEQ ID NO 13938; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for

XX measuring human gene expression in a sample derived from human fetal

XX liver. The single exon nucleic acid probes may be used for predicting,

XX measuring and displaying gene expression in samples derived from human

XX fetal liver. The present sequence is a single exon nucleic acid

XX probe of the invention.

XX Note: The sequence data for this patent did not form part of the

XX printed specification, but was obtained in electronic format directly

XX from WIPO at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 1785 BP; 293 A; 622 C; 480 G; 390 T; 0 other;

Query Match 2.5%; Score 47; DB 22; Length 1785;

Best Local Similarity 49.4%; Pred. No. 0.089;

Matches 122; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 1062 CTTCGGCTTCAGAGCAAGACCTCGAGATAGAGCGCTGTACTCAGCGCGAGGTTAA 1121

DB 510 CATGCCCTTGAATTGGCCCTACCTGAGAGTCTCTGGCCAGACCTTCATCACTGCTGT 569

QY 1122 AGAGGAG 1181

DB 570 CGGCCACCTGTGTCTCTCAACGACGACGCTCCGCTGCTCTATGTCTACCTGCTCTA 629

QY 1182 CCGCTTCGTACGAGATGAGCTGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1241

DB 630 TCTCTTCTTCGAGTGCACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 689

QY 1242 CCGCGCGGAG 1301

DB 690 CTACCTGTGTGTCTCTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 749

QY 1302 CTTCGAG 1308

DB 750 CCTGGGG 756

Search completed: July 15, 2003, 14:01:32
Job time : 436 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2003, 11:46:09 ; Search time 4928 Seconds

(Without alignments)
11037.570 Million cell updates/sec

Title: US-09-966-803-1

Perfect score: 1869

Sequence: 1 ATGACCGCATCGAATGGA.....TCAAGGGGAATGGAACTGA 1869

Scoring table: IDENTITY_NUC

Gapop 10.0 / Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

GenBank:
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pin:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1869	100.0	1869	6	AR086952 Sequence
2	1869	100.0	1869	6	AR095647 Sequence
3	1869	100.0	1869	6	AR136115 Sequence
4	1078.2	57.7	304290	1	ARJ42827 Pyrococcus
5	1078.2	57.7	349980	6	CNSPAX05
6	955	51.1	257000	1	AX041921 Sequence
7	825.2	44.2	13354	1	AP000002 Pyrococcus
8	522	27.9	11370	1	AE010156 Pyrococcus
9	204.8	11.0	247695	1	AE013111 Thermococcus
10	166	8.9	10256	1	AE000064 Aeropyrum
11	150.8	8.1	265118	1	AE001879 Deinococcus
12	150.8	8.1	265118	6	CNSPAX06
13	150.2	8.0	217420	1	AX041922 Sequence
14	149.8	8.0	1971	6	BSUB0017
15	140.8	7.5	1958	6	AX006271 Bacillus su
16	132.2	7.1	11233	1	AX006271 Sequence
17	131	7.0	251700	1	AX431809 Sequence
18	120.4	6.4	1896	6	AE009798 Pyrobacul
19	120.4	6.4	1896	6	AR116209
20	120.8	6.4	233000	1	E16634 Pyrococcus
21	102.8	5.5	12882	1	AP000003 Pyrococcus
22	98.6	5.3	9825	1	AE013176 Thermococcus
23	96.8	5.2	139203	2	AE005114 Halobacter
24	92.6	5.0	11388	1	AC068311 Homo sapi
25	84.6	4.5	2642	1	AE011750 Xanthomon
26	79.8	4.3	12098	1	AF508804 Xanthomon
27	72	3.9	182440	1	AE012209 Xanthomon
28	72	3.9	316900	1	AP000994 TACID3
29	68.4	3.7	17112	1	AE004997 Thermopla
30	66.2	3.5	477	6	AX436451 Sequence
31	61.8	3.3	3808	1	AF068743 Haloferrax
32	58.6	3.1	12358	1	AE006863 Sulfolobu
33	58.4	3.1	1836	6	AX078507 Sequence
34	57.8	3.1	7459	1	SC8A2 Streptomy
35	57.8	3.1	1836	1	SC8A2 Streptomy
36	56.2	3.0	125020	9	AF429315 Homo sapi
37	55	2.9	11105	1	AE004963 Pseudomon
38	54.4	2.9	39446	1	SCER7 Streptomy
39	53.4	2.9	38634	1	SC2665 Streptomy
40	52.8	2.8	3217	1	SGSECA Streptomy
41	52.2	2.8	12212	1	AE005872 Caulobact
42	51.4	2.8	268050	1	AP000983 Streptomy
43	50.6	2.7	26423	1	SC8G12 Streptomy
44	50.4	2.7	125020	9	AF429315 Homo sapi
45	50.4	2.7	167446	2	OSJN0085 Oryza sat

ALIGNMENTS

RESULT 1
LOCUS AR086952 1869 bp DNA
DEFINITION Sequence 1 from patent US 5985646.
ACCESSION AR086952
VERSION AR086952.1 GI:10013718
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1869)
AUTHORS Murphy,D., Reid,J. and Robertson,D.
TITLE Amidase
JOURNAL Patent: US 5985646-A 1 16-NOV-1999;
FEATURES Location/Qualifiers

source 1..1869
 /organism="unknown"
 BASE COUNT 476 a 486 c 567 g 340 t
 ORIGIN

Query Match 100.0%; Score 1869; DB 6; Length 1869;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1869; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGGGCAATCGAATGGAACACGAGACCTTTTCTAAGTTCGCTTACCTGGGACCGG 60
 Db 1 ATGACGGGCAATCGAATGGAACACGAGACCTTTTCTAAGTTCGCTTACCTGGGACCGG 60
 QY 61 AGGATACGGGCAATCGAATGGAACACGAGACCTTTTCTAAGTTCGCTTACCTGGGACCGG 120
 Db 61 AGGATACGGGCAATCGAATGGAACACGAGACCTTTTCTAAGTTCGCTTACCTGGGACCGG 120
 QY 121 TACGAGACGAGGCTTGTGTTGAAGACCTTGAACGGGCTTCAGAGGCTTCATGAGAAC 180
 Db 121 TACGAGACGAGGCTTGTGTTGAAGACCTTGAACGGGCTTCAGAGGCTTCATGAGAAC 180
 QY 181 GCTCAATGCCGAGGATTTGCCAGACGGCAGAAAGCTCGCTTACCTGCTTAAACGAG 240
 Db 181 GCTCAATGCCGAGGATTTGCCAGACGGCAGAAAGCTCGCTTACCTGCTTAAACGAG 240
 QY 241 GAGAAAGAGAGACGAGATATGGGTGGCCGATATCCAGACCCCTGAGGCGCAAGAAAGTC 300
 Db 241 GAGAAAGAGAGACGAGATATGGGTGGCCGATATCCAGACCCCTGAGGCGCAAGAAAGTC 300
 QY 301 CTCTCAATAAAAAGTCCGCTCGATCGATGAGACGATTCAGAGAGACTTCTAGTT 360
 Db 301 CTCTCAATAAAAAGTCCGCTCGATCGATGAGACGATTCAGAGAGACTTCTAGTT 360
 QY 361 GTGGGCTTCAAG 420
 Db 361 GTGGGCTTCAAG 420
 QY 421 TTGACAAATATGGGATTTCTTGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
 Db 421 TTGACAAATATGGGATTTCTTGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
 QY 481 GCCGAGAGAGATATCGAGCAGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
 Db 481 GCCGAGAGAGATATCGAGCAGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
 QY 541 GATGCGATATGTTGGAAGTCCGCGACGCGAGGGGGAGCAAGCTTCCCTGTTAAAGTC 600
 Db 541 GATGCGATATGTTGGAAGTCCGCGACGCGAGGGGGAGCAAGCTTCCCTGTTAAAGTC 600
 QY 601 TACGACATATGCTATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
 Db 601 TACGACATATGCTATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
 QY 661 GAGGCGGTGACTCCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
 Db 661 GAGGCGGTGACTCCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
 QY 721 ATGACGAG 780
 Db 721 ATGACGAG 780
 QY 781 CCGCTCGAGCTCTGGAAGCCAGCTCAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 840
 Db 781 CCGCTCGAGCTCTGGAAGCCAGCTCAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 840
 QY 841 GCGGCGAG 900
 Db 841 GCGGCGAG 900
 QY 901 CACTGAGATTAAGGAGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
 Db 901 CACTGAGATTAAGGAGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960

QY 961 ACGAGATATGAGGAGAGCTTACCTTACAGAGGCGAGAGAGAGAGAGAGAGAGAGAGAG 1020
 Db 961 ACGAGATATGAGGAGAGCTTACCTTACAGAGGCGAGAGAGAGAGAGAGAGAGAGAGAG 1020
 QY 1021 GGGCGGATATTCAG 1080
 Db 1021 GGGCGGATATTCAG 1080
 QY 1081 GACCTCGAGATATGAGGAG 1140
 Db 1081 GACCTCGAGATATGAGGAG 1140
 QY 1141 ATAGCTTCGTCACGAGGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
 Db 1141 ATAGCTTCGTCACGAGGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
 QY 1201 CAGCTGATATGAGGAG 1260
 Db 1201 CAGCTGATATGAGGAG 1260
 QY 1261 TATAGCGAG 1320
 Db 1261 TATAGCGAG 1320
 QY 1321 ATATGAG 1380
 Db 1321 ATATGAG 1380
 QY 1381 GGAATTAACGGGCAATAGCTTACGAGGCGCTTATGAGCAACATGAGGCTTACCTGAGAG 1440
 Db 1381 GGAATTAACGGGCAATAGCTTACGAGGCGCTTATGAGCAACATGAGGCTTACCTGAGAG 1440
 QY 1441 CTCTTCAAG 1500
 Db 1441 CTCTTCAAG 1500
 QY 1501 TCAGACATATGAGGCTTCTGTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
 Db 1501 TCAGACATATGAGGCTTCTGTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
 QY 1561 AACTTCAAG 1620
 Db 1561 AACTTCAAG 1620
 QY 1621 ATCCAGCTGTTGAG 1680
 Db 1621 ATCCAGCTGTTGAG 1680
 QY 1681 CTCAAG 1740
 Db 1681 CTCAAG 1740
 QY 1741 AGCGTCCGCGAG 1800
 Db 1741 AGCGTCCGCGAG 1800
 QY 1801 GAGGCGAAGCTCAAG 1860
 Db 1801 GAGGCGAAGCTCAAG 1860
 QY 1861 GGAAGCTGA 1869
 Db 1861 GGAAGCTGA 1869

RESULT 2
 AR095647 1869 bp DNA linear PAT 08-SEP-2000
 LOCUS AR095647
 DEFINITION Sequence 1 from patent us 6004796.
 ACCESSION AR095647
 VERSION AR095647.1 GI:10023708
 KEYWORDS
 SOURCE unknown.

ORGANISM Unknown.
 Unclassified.
 REFERENCE 1 (bases 1 to 1869)
 AUTHORS Murphy, D., Reid, J. and Robertson, D.
 TITLE Amidase
 JOURNAL Patent: US 6004796-A 1 21-DEC-1999;
 FEATURES Location/Qualifiers
 source 1.1869
 BASE COUNT 476 a 486 c 567 g 340 t
 ORIGIN

Query Match 100.0%; Score 1869; DB 6; Length 1869;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1869; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGACCGGCATCGATGAGACACGACCTTTTCTAAGTTCGCTTACCTGGGCGACCG 60
 1 ATGACCGGCATCGATGAGACACGACCTTTTCTAAGTTCGCTTACCTGGGCGACCG 60
 61 AGATACCGGGAACCTTAATCGCGTACACCTGACGAAGGCCAATGAGCAACAAG 120
 61 AGATACCGGGAACCTTAATCGCGTACACCTGACGAAGGCCAATGAGCAACAAG 120
 121 TACGAGACGACGGTTGTTGTAAGACCTTGAACGGGCTCAAGGCGCTTATGAGAAC 180
 121 TACGAGACGACGGTTGTTGTAAGACCTTGAACGGGCTCAAGGCGCTTATGAGAAC 180
 181 GCGTCAATGCCGAGATTTCCGACAGCGCAAAAGCTCGCTTACCTGCTTTAAGCAG 240
 181 GCGTCAATGCCGAGATTTCCGACAGCGCAAAAGCTCGCTTACCTGCTTTAAGCAG 240
 181 GCGTCAATGCCGAGATTTCCGACAGCGCAAAAGCTCGCTTACCTGCTTTAAGCAG 240
 241 GAGAAGAAAGAGACCGGATATGCGGTGCGCGATATCCAGACCTGAGCGCAAGAAATC 300
 241 GAGAAGAAAGAGACCGGATATGCGGTGCGCGATATCCAGACCTGAGCGCAAGAAATC 300
 241 GAGAAGAAAGAGACCGGATATGCGGTGCGCGATATCCAGACCTGAGCGCAAGAAATC 300
 301 CTCTCACTAATAAAAGTCCGCTGATGAGTGAAGCAAGCAATGCAAGAGACTTAACTT 360
 301 CTCTCACTAATAAAAGTCCGCTGATGAGTGAAGCAAGCAATGCAAGAGACTTAACTT 360
 361 GTGCGCTTCAAGAGAGAGAGATGAGACTTCTCTTTGACGACGAGCTCCGCTGG 420
 361 GTGCGCTTCAAGAGAGAGAGATGAGACTTCTCTTTGACGACGAGCTCCGCTGG 420
 421 TTGCAACAATATGGGATTTCTTTGATGAGAGAAAGACGCTTCTGGGTTCTTACACTGAG 480
 421 TTGCAACAATATGGGATTTCTTTGATGAGAGAAAGACGCTTCTGGGTTCTTACACTGAG 480
 421 TTGCAACAATATGGGATTTCTTTGATGAGAGAAAGACGCTTCTGGGTTCTTACACTGAG 480
 481 GCCGAGAGATATATCGAGAGTTGAGAGACCGAGGTTTTCAGTGGGCTTGGCAGCGC 540
 481 GCCGAGAGATATATCGAGAGTTGAGAGACCGAGGTTTTCAGTGGGCTTGGCAGCGC 540
 481 GCCGAGAGATATATCGAGAGTTGAGAGACCGAGGTTTTCAGTGGGCTTGGCAGCGC 540
 541 GATGCGATATGTTGTAAGCTCCGACACCGGAGGAGCAAGCTTCCCTGTTCAAGTTC 600
 541 GATGCGATATGTTGTAAGCTCCGACACCGGAGGAGCAAGCTTCCCTGTTCAAGTTC 600
 541 GATGCGATATGTTGTAAGCTCCGACACCGGAGGAGCAAGCTTCCCTGTTCAAGTTC 600
 601 TAGGACATATGCTTATGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 660
 601 TAGGACATATGCTTATGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 660
 601 TAGGACATATGCTTATGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 660
 661 GAGGCGGTTGACTCCGAGGAAAGAGATATCTCTGAGGGGCAAGAAAAGGCGGTTTC 720
 661 GAGGCGGTTGACTCCGAGGAAAGAGATATCTCTGAGGGGCAAGAAAAGGCGGTTTC 720
 661 GAGGCGGTTGACTCCGAGGAAAGAGATATCTCTGAGGGGCAAGAAAAGGCGGTTTC 720
 721 ATCAGGAGACGACTGCTGTATCTCTGAGGAGCGGAGCTTAAACGAGCTTACGAGGCG 780
 721 ATCAGGAGACGACTGCTGTATCTCTGAGGAGCGGAGCTTAAACGAGCTTACGAGGCG 780
 721 ATCAGGAGACGACTGCTGTATCTCTGAGGAGCGGAGCTTAAACGAGCTTACGAGGCG 780
 781 CCGCTGACGCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
 781 CCGCTGACGCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
 781 CCGCTGACGCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
 841 GCGGCGAGGCTAAACCTCTGCTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900

|||||
 841 GCGGCGAGGCTAAACCTCTGCTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
 901 CACTGAGATTTAGGGCTTACGCTGACGATGCGCAAGATGCTCTCTATCATGACGCC 960
 901 CACTGAGATTTAGGGCTTACGCTGACGATGCGCAAGATGCTCTCTATCATGACGCC 960
 901 CACTGAGATTTAGGGCTTACGCTGACGATGCGCAAGATGCTCTCTATCATGACGCC 960
 961 ACGAGATATGAGGAGCTTACCTTACGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
 961 ACGAGATATGAGGAGCTTACCTTACGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
 961 ACGAGATATGAGGAGCTTACCTTACGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
 1021 GGGCGGATATTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
 1021 GGGCGGATATTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
 1021 GGGCGGATATTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
 1081 GAGCTGAGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
 1081 GAGCTGAGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
 1081 GAGCTGAGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
 1141 ATAGCTTCTGTCACGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
 1141 ATAGCTTCTGTCACGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
 1141 ATAGCTTCTGTCACGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
 1201 CAGCTGATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
 1201 CAGCTGATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
 1201 CAGCTGATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
 1261 TATAGCGAAGACTTGGCGCTCGCTCGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320
 1261 TATAGCGAAGACTTGGCGCTCGCTCGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320
 1261 TATAGCGAAGACTTGGCGCTCGCTCGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320
 1321 ATATGAAAGGCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
 1321 ATATGAAAGGCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
 1321 ATATGAAAGGCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
 1381 GGAATTAAGGCGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
 1381 GGAATTAAGGCGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
 1381 GGAATTAAGGCGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
 1441 CTCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500
 1441 CTCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500
 1441 CTCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500
 1501 TCGGACATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
 1501 TCGGACATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
 1501 TCGGACATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
 1561 AACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
 1561 AACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
 1561 AACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
 1621 ATCCACTGCTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680
 1621 ATCCACTGCTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680
 1621 ATCCACTGCTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680
 1681 CTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740
 1681 CTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740
 1681 CTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740
 1741 AGGCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800
 1741 AGGCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800
 1741 AGGCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800
 1801 GAGGCGAAGCTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860
 1801 GAGGCGAAGCTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860
 1801 GAGGCGAAGCTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860
 1861 GGAAGCTGA 1869
 1861 GGAAGCTGA 1869
 1861 GGAAGCTGA 1869

RESULT 3

ARI36115 ARI36115 .1869 bp DNA linear PAT 16-JUN-2001
LOCUS Sequence 1 from patent US 6136583.
DEFINITION ARI36115
ACCESSION ARI36115
VERSION ARI36115.1 GI:14476787
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1869)
AUTHORS Murphy, D., Reid, J. and Robertson, D.
TITLE AMIDASE
JOURNAL Patent: US 6136583-A 1 24-OCT-2000;
FEATURES
source Location/Qualifiers
BASE COUNT 476 a 486 c 567 g 340 t
ORIGIN
Query Match 100.0%; Score 1869; DB 6; Length 1869;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1869; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGACCGGCATCGAATGGAACACAGACCTTTCTAAGTTGCGCTACTGGGGACCCG 60
DB 1 ATGACCGGCATCGAATGGAACACAGACCTTTCTAAGTTGCGCTACTGGGGACCCG 60
QY 61 AGGATACGGGGAAACTTAATCGCGCTACACCTGACGAAGGCAATGAAGACACAG 120
DB 61 AGGATACGGGGAAACTTAATCGCGCTACACCTGACGAAGGCAATGAAGACACAG 120
QY 121 TACAGAGACCGGTGTGTGTGAAGACCTTGAACGGGCTCAAGCCCTCATCGAAGAC 180
DB 121 TACAGAGACCGGTGTGTGTGAAGACCTTGAACGGGCTCAAGCCCTCATCGAAGAC 180
QY 181 GCCCTAATGCCGAGGATTTGCCAGACGGCAGAAAGCTCGCTTCACTGCTTTAAGAG 240
DB 181 GCCCTAATGCCGAGGATTTGCCAGACGGCAGAAAGCTCGCTTCACTGCTTTAAGAG 240
QY 241 GAGAAAGAGAGACCGAGATTTGGGTGGCGGATTTCCAGACCTTGAGCGCCAAAGAGTC 300
DB 241 GAGAAAGAGAGACCGAGATTTGGGTGGCGGATTTCCAGACCTTGAGCGCCAAAGAGTC 300
QY 301 CTCTCACTAAAAAGTCGCTCGATGAGTGAAGCAGATTCAGAGAGACTTTAGTT 360
DB 301 CTCTCACTAAAAAGTCGCTCGATGAGTGAAGCAGATTCAGAGAGACTTTAGTT 360
QY 361 GTCGGCTTCAAGAGAGAGAGAGATGAGAGACTTCTTTGAAGCAGAGCTCCGCTGG 420
DB 361 GTCGGCTTCAAGAGAGAGAGAGATGAGAGACTTCTTTGAAGCAGAGCTCCGCTGG 420
QY 421 TTGCACAATATGAGATTTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 421 TTGCACAATATGAGATTTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
QY 481 GCCGAGAGAGATTAATCGAGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
DB 481 GCCGAGAGAGATTAATCGAGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
QY 541 GATCGATAGTTGTTGAAGCTCCGACCGGAGGAGGAGAGAGAGAGAGAGAGAGAGAG 600
DB 541 GATCGATAGTTGTTGAAGCTCCGACCGGAGGAGGAGAGAGAGAGAGAGAGAGAGAG 600
QY 601 TACGACATATGATTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 601 TACGACATATGATTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 661 GAGCGGTGACTCGAG 720
DB 661 GAGCGGTGACTCGAG 720
QY 721 ATCAGCAGACGAGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
DB 721 ATCAGCAGACGAGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780

DB 721 ATCAGCAGACGAGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
QY 781 CCGCTCGACGCTCTGGAG 840
DB 781 CCGCTCGACGCTCTGGAG 840
QY 841 GCGGCGAGAGGTAACCTCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB 841 GCGGCGAGAGGTAACCTCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 901 CACTGATATACGGGCTTGAAG 960
DB 901 CACTGATATACGGGCTTGAAG 960
QY 961 ACAGAGATAGGCGAGCTCTACCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
DB 961 ACAGAGATAGGCGAGCTCTACCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
QY 1021 GGGCGATATTCAG 1080
DB 1021 GGGCGATATTCAG 1080
QY 1081 GACCTCGAGATAG 1140
DB 1081 GACCTCGAGATAG 1140
QY 1141 ATAGCTCTGCTCAAGCGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1141 ATAGCTCTGCTCAAGCGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1201 CAGCTGATAGGAG 1260
DB 1201 CAGCTGATAGGAG 1260
QY 1261 TATAGCAG 1320
DB 1261 TATAGCAG 1320
QY 1321 ATATGAG 1380
DB 1321 ATATGAG 1380
QY 1381 GGAATTAAG 1440
DB 1381 GGAATTAAG 1440
QY 1441 CTCTTCAAG 1500
DB 1441 CTCTTCAAG 1500
QY 1501 TCGGACATAGGAG 1560
DB 1501 TCGGACATAGGAG 1560
QY 1561 AACTTCAG 1620
DB 1561 AACTTCAG 1620
QY 1621 ATCCATCTGCTTGAAG 1680
DB 1621 ATCCATCTGCTTGAAG 1680
QY 1681 CTCAAG 1740
DB 1681 CTCAAG 1740
QY 1741 AGCGTCGCGAG 1800
DB 1741 AGCGTCGCGAG 1800
QY 1801 GAGCGCAAGCTCAAG 1860
DB 1801 GAGCGCAAGCTCAAG 1860

Query Match	Best Local Similarity	Matches 1405; Conservative	57.7%; 74.9%; 0;	Score 1078.2; Pred. No. 2,6e-201; Mismatches 458;	DB 6; Indels 12; Gaps 4;	Length 349980; from 0.000.001 to 1.765.118; length: 265.118"
1	ATGACCGGCATTCGAAATGGAACCAACGACACCTTTTCTAAGTTGGCTTACCTGGGGACCCG	60				
206185	ATGACCTCTATCGAATGGAATGGAAGACGTTTACAAAGTTGGCTTACCTGAAAGGACCC	206128				
61	AGATATCGGGGAAACTTAATCGGCTACACCTCGACGAAGGCCAATGAAGCAACAAG	120				
206125	AGATATGAGGGGAAAGCAACATAGCTGATCTTACCAAGGCCAATCTTGAACACACAG	206066				
121	TACGAGACGACGCTTGTGTTGAAGCCTTGAACGGGCTCAAGGCTCATCGAAGC	180				
206065	TACAGAGCAGATGATGTTCTCGAAGACCTCGAAGATGGAGACAGAGGATCATAGAAAG	206006				
181	GCCCTAATGCGAGGATTTGCGACAGCGCAAGAAAGCTGCGCTTACCTGCTTTAAAGAG	240				
206005	GCTTCAATGCTCTAGATTTCTCCAGATGGAAGAAAGAAAGATACCTTATATAGGTTACAGAG	205946				
241	GAGAAAGAGAGACCGAGATATGATGGTGGCCGATATCCAGACCTCGAGCCCAAGAAATC	300				
205945	GAGAAAGAAATGTTCTCGAATATGGTGGCGGACATGAAAGACCTTAAAGCCCAAGAAATGTT	205886				
301	CTCTCAACTTAAAAAGCTGCGCTCGATGACGATGGAAGCAAGCAAGTCAAGGAGACCTTAGTT	360				
205885	CTCGAGGCCAAGACATTAAGTCTCTGGAATGGAATGACGACTCAAGAGGTTGTAGTT	205828				
361	GTCGGCTTCAAGAGGAGGAGACGATGAAGACTTCTGTTTGAACGACGACGTCGCCGTGG	420				
205825	ATAGGCTTCAAGAGGAGGAGGACGATGAAGATTTTCATATTCGAGAGTGAAGTTCACAGCTTGG	205766				
421	TTGCAACAATTTGGGATTTCTTTTATGATGAGAGAAAGACAGACTTCTGGGTTCTTGACACTGAG	480				
205765	TTTGACACAACATGGGCTTCTTTCGACGGCGGAGAAAGACGCTTCTGGATATGATGAATAGAG	205706				
481	GCGAGAGAGATTAATCGAGAGATTCGAGAGAGCCGAGGTTTTCGAGTGGCTTCGACAGGC	540				
205705	GCGAGAGAGGTTTATAGAGCAAGTTTCGAGAAAGCCAAAGTTTTCGAGCGAGATTTGGCACGGG	205648				
541	GATGCGATTAATTGTGAACGTCGCCGACGCGAGGGGAGACGCTGCTGTTCAGTTTC	600				
205645	GATTCGATTAATTGTGAACGTCGCCGTCAGC---GGACACCAATTAACAGGTAAGTTCAAGTTAC	205589				
601	TACGACATTAATCTTATGGAAGAGCGGGGAGGAAGAAAGCACTTCCTCGAGAGGCTCTCTTC	660				
205588	TGGAACTCTTAATCTCTGGAAGGTGAGGAGGAAGAAAGCACTTCTCGAGAGGCTCTCTTC	205529				
661	GAGCGGTTGATCTCCACGAGAAAGAAATATCTCGAGGGGCAAGAAAAAAGCGGTTTC	720				
205528	CAGGCTATTAATCTCTGAGCGAAAGAAATATCTTACTCTACGGAAGCCCGAGAAAGAAATAC	205469				
721	ATCAGAGACGACGATGCTGTATCTCTGTGGAGCGGCAAGTTAA---ACGACTTACGAG	777				
205468	ATGAGGAGACGATTAAGCTCTCAACATCTACGACGCGGAGGTTAAGGGAATAAATGAGATTCG	205409				
778	GCGCCGCTGATCTCTGGGAGAGCAAGCTTCAGAGGAAGAAAGTCTACTTCTTCATCTCA	837				
205408	ATGACAGAGGAAGCTGCTACAGCTTAAGATCAAGAGATGGGAAGGTGTACTTCTCACACTTCAC	205349				
838	GATGCGGAGGAGGCTTAACCTTGCGCTCTTGGAGCGGGAAGGCGGACGCTGTGTTACCGGC	897				


```

Db      205348 GAGGAGGAGGAGGCTTACCTATATCTCTGGGATGGGGAATATCAAGAGATAGCTAAGGCG 205289
QY      898 GACCACTGGATTATACGGGCTTGAGCTCAGCGATGCGATTCCTCTCCATCATGAGC 957
Db      205288 AAGCACTGGATTATATGGGCTTACAGCT---TGACAGAAAGAGTGTCTACTCTCAAGAACT 205232
QY      958 GCCAGAGAGATAGGAGGCTCTACCTCTACGACGGCGAGCTCAAAAGAGTCAAGCAATAC 1017
Db      205231 GCCACTAGGTTAGGAGGCTCTACATCTGGAGACGGCGAGAGAGCAACTTAACGGATTAC 205172
QY      1018 AACGGGCCGATATATAGAGAGCTCAAGAGCTTCGAGCGAGCGAGCTTCGCTCTCAAGAGC 1077
Db      205171 AACGCTCTATATATAGAGAGCTCAAAACCTTCGAGCGCTTAAGCACTTCAAGTTCAGAGC 205112
QY      1078 AAAAGCTCTGAGATAGAGGCTGTGATACCTCAGCGCGAGGTTAAAG--GAGAAGGCC 1134
Db      205111 ATCGACTTGGATAGAGGCTGTGATACATTAAGCCCGAGCTCAAGAGGCGGAGAAAGCT 205052
QY      1135 CGGCTGATATGTTCTGTCACGCGCGCGGAGGAGGAGTATAGAGACCCGCTTCTGCTAC 1194
Db      205051 CGGCTGATATGTTCTGTCACGCGCGCGGAGGAGTATAGAGACCCGCTTCTGCTAC 204992
QY      1195 GAGATGCAAGCTGATGGCGAGCAAGGCTACTACTGCTGCTGCTGTAACCCGCGCGAGC 1254
Db      204991 GAGATGCAAGCTTATGGCGAGCAAGGCTACTACTGCTGCTGCTGTAACCCGCGAGC 204932
QY      1255 GACGCTATAGCGAAGACTTCGCGCTCGCGCTCTGAGAGAGAGCTGGCTTGAAGACTTT 1314
Db      204931 AACGCTTACAGGAGAGACTTCTGCTGCTGCTGCTGAGAGAGAGCTGGCTTGAAGACTTT 204872
QY      1315 GAGGATATATAGAGGCTCTGAGAGGCTTCTCAAGCTGAGAGCGAGCGAGCGAGGAG 1374
Db      204871 CAGGATATATCTAATAGAGATAGAGAGGTTCTTCAAGCTTGAACCGGAGCGAGCGAGGAG 204812
QY      1375 CGCGTTGGAATAGCGGCAATAAGCTACGCGGCTTCATGACCACTGGCGCTTGAAGTAC 1434
Db      204811 AAGGTTGGAATAGCGGCAATAAGCTACGAGGCTTCATGACCACTGGCGCTTGAAGTAC 204752
QY      1435 AACGACCTCTTCAAGGAGAGATTAAGCGAGAGCGCATTAAGTACTGGCTCAACGCTAC 1494
Db      204751 AACGACCTTCAAGGAGAGATTAAGCGAGAGCGCATTAAGTACTGGCTCAACGCTAC 204692
QY      1495 GCGCTTCTGAGATAGGCGCTGTGTAAGAGCTGAGAGCTCATGCGGCAATCCGTTAGAG 1554
Db      204691 GCGCTTCTGAGATAGGCGCTGTGTAAGAGCTGAGAGCTCATGCGGCAATCCGTTAGAG 204632
QY      1555 AACGAGACTTCAAGAGAGCTCAAGCGCGCTGTCTACGCTCAGAGAGCGTGAAGCGCGGATA 1614
Db      204631 AACGAGACTTCAAGAGAGCTCAAGCGCGCTGTCTACGCTCAGAGAGCGTGAAGCGCTG 204572
QY      1615 CTCCTATATCACTCGCTTGAAGAGCTACCGCTGTCCGCTGACCAAGGCTTATGTTCTAC 1674
Db      204571 TTGCTATATCAAGCGCTTGAAGAGCTTATGCTGTCTCAATGATGAGTGAAGTGTCTAC 204512
QY      1675 AACGCTCTCAAGAGATGGCGAGAGAGCTTCAATAGCATATTCATTAAGCGCGCGCGC 1734
Db      204511 CACGCTCTCAAGAGATTTAGCGAAAGATTACATAGCATATTCATTAAGCGCGCGAGCTCAT 204452
QY      1735 GCCCAGAGCGTCCGCGGAGAGCGCGAGGACAGCGCGAGCGCTTACAGCTTCTTCAATAG 1794
Db      204451 GGACCTCTATATAGGAGAGCGCGAGAGCGCGAGGACAGCGAGAGCTTACAGCTTCTTCAATAG 204392
QY      1795 TTCTTCTGAGAGAGCTCAAGAGATTAAGAGAGAGGCTTGAAGTGAAGAGTACTTAAAG 1854
Db      204391 TTCTTCTGAGAGAGCTCAAGAGATTAAGAGAGAGGCTTGAAGTGAAGAGTACTTAAAG 204332
QY      1855 GGGAAATGGGAACTGA 1869
Db      204331 GGTGAGAAACAAGTGA 204317

```

RESULT 6

```

AP000002
LOCUS      257000 bp      DNA      linear      BCT 06-APR-2000
DEFINITION Pyrococcus horikoshii OT3 genomic DNA, 287001-544000 nt. position (2/7).
ACCESSION AP000002 AB009475 AB009476 AB009477 AB009478 AB009479 AB009480
AB009481 AB009482 AB009483 AB009484 BA000001
VERSION    AP000002.1 GI:3236129
KEYWORDS   Pyrococcus horikoshii (strain:OT3) DNA.
SOURCE     Pyrococcus horikoshii
ORGANISM   Pyrococcus horikoshii
REFERENCE  1 (sites)
            Kawarabayashi, Y., Sawada, M., Horikawa, H., Halkawa, Y., Hino, Y.,
            Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y.,
            Sakai, M., Ogura, K., Otsuka, R., Nakazawa, H., Takamiya, M., Ohikubo, Y.,
            Funahashi, T., Tanaka, T., Kudo, Y., Yamazaki, J., Kishida, N.,
            Onouchi, A., Aoki, K., Nakamura, Y., Robb, T. F., Horikoshi, K.,
            Masuchi, Y., Shizuya, H. and Kikuchi, H.
            Complete sequence and gene organization of the genome of a
            hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3
            DNA Res. 5 (2), 55-76 (1998)
            98344137
            2 (bases 1 to 257000)
            Tanaka, T., Kawarabayashi, Y. and Kikuchi, H.
            Direct Submission
            Submitted (11-JUN-1998) Yutaka Kawarabayashi, National Institute of
            Biotechnology and Evaluation, Biotechnology Center: 2Chome 49-10
            Nishihara, Shibuya-Ku, Tokyo 151-0066, Japan
            (E-mail: genome03@ente.go.jp, Tel: +81-3-3481-8951,
            Fax: +81-3-3481-8424)
            On or before Jul 28, 1998 this sequence version replaced
            91:3130553, 91:3130577, 91:3130627, 91:3130656, 91:3130685,
            91:3130705, 91:3130737, 91:3130783, 91:3130824, 91:3130828,
            AB009475-AB009484: submitted (10-DEC-1997)
            Kawarabayashi, Y. is officially affiliated with the National
            Institute of Biotechnology and Human-Technology, Tsukuba, Ibaraki
            305-0046, Japan.
            Robb, T. F. is at the Center of Marine Biotechnology, University of
            Maryland, Baltimore, MD, USA.
            Horikoshi, K. is at the Japan Marine Science and Technology Center,
            Yokosuka, Kanagawa 237-0061, Japan.
            Masuchi, Y. is at the University of Tokyo, Meguro, Tokyo 153-0041,
            Japan.
            Shizuya, H. is at the California Institute of Technology, Pasadena,
            CA, USA.
            The other authors are at the National Institute of Technology and
            Evaluation, 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan.
            All the sequence with length 100 codons or more between ATG or GTG
            and stop codon are defined as CDS
            Homology analysis is performed by Smith-Waterman algorithm against
            Genbank and Genpept release 103; EMBL release 52.0; SwissProt
            release 34.0; PIR-Protein release 54.0; and OML release 29.5.
            E-mail address for comments and questions: genome03@ente.go.jp
            Information are available at W.W.W. site of Biotechnology Center,
            URL: http://www.dlo.ente.go.jp/.

```

FEATURES

```

source     1..257000
            /organism="Pyrococcus horikoshii"
            /strain="OT3"
            /db_xref="taxon:53953"
            /map="287001-544000 nt. position"
            /length=257000
            /gene="PH0318"
            /gene="PH0318"
            /gene="PH0318"
            /note="similar to PIR:E64485 percent identity: 55.181 in
            403aa; PIR:S74842 percent identity: 49.869 in 397aa;
            Swiss Prot:P33221 percent identity: 50.773 in 404aa"
            /codon_start=1
            /transl_table=11
            /product="433aa long hypothetical"

```


TITLE	The complete sequence of the <i>Pyrococcus furiosus</i> genome	
	Unpublished	
REFERENCE	4 (bases 1 to 13354)	
AUTHORS	Weiss, R.B.	
JOURNAL	Direct Submission	
FEATURES	Submitted (12-FEB-2002) Human Genetics, University of Utah, 20	
	South 2030 East, Salt Lake City, UT 84112, USA	
SOURCE	Location/Qualifiers	
	1. 13354	
	/organism="Pyrococcus furiosus DSM 3638"	
	/strain="DSM 3638"	
	/db_xref="taxon:186497"	
gene	94. 1113	
CDS	/gene="PF0308"	
	94. 1113	
	/note="Function Code: 9.3 Purines, Pyrimidines, Nucleosides and Nucleotides: Purine ribonucleotide biosynthesis: (pura)"	
	/codon_start=1	
	/transl_table=11	
	/product="adenylosuccinate synthetase"	
	/protein_id="AAL80432.1"	
	/db_xref="GI:18892258"	
	/translation="MPSVIVGGGMDGEGKSTIAYIALHDEPITARGVGYNAGHS VVINGKRYAVROLPTGPMQKRALINGAGVLDPEVFTLEHLKDFNVAREVGLDIR CAIEPRHKMKNDLHGISTGTSGGCPANADRMVAKAKDKIETELPLTDVA AEVNDLDEGALVLEGTQGFGLSYLYGTPYVTSKDTASAVASDVGIGPTVDVI VPKSPPTRGVAGPPTPEMPEBADRGLILEYGVTVGRRRVGMFDEPMARYARING ATMLAVTMDKRYKEAFGTDYDKLPKRAKEFIEIEBEVGVGLIKTGPHEHIIID LRENI"	
tRNA	complement(1174. 1251)	
	/product="tRNA-Glu"	
gene	/note="tRNA Glu anticodon TTC"	
CDS	complement(1303. 2580)	
	/gene="PF0309"	
	complement(1303. 2580)	
	/gene="PF0309"	
	/note="Function Code: 14.6 Transport and Binding Proteins: Other"	
	/codon_start=1	
	/transl_table=11	
	/product="ABC transporter (membrane protein)"	
	/protein_id="AAL80433.1"	
	/db_xref="GI:18892259"	
	/translation="MSDFVVLVTKKLDLRDKGLIFGIIVPLIITPALGQMLQVGF EDAQETKVVVNFDEGRYGNILIKALSNPVTVKIEAKTIDEALKRAQEEYNNI VIIPKNSQAIENSKASVEIYIGIVGISGSMREASSEGINAVISILNEYIAKIE RRIEGDPEALKIDAKSYTIKRIYVEIPSVVSNIIASQSPSMELVFIEMIMIAO MSAGSMAMKENTLETLLTLPYKRITIVAGVGAIIIGIIAISMGKNTYGLK TSSAOVSTLEELGKITPLGATLFIITFIVTVLSRAMLAIVAEQDKANTAVY AGIMPLAFPTFILMPADVNTLPTVLKLLGLIPFSPIILASRAMLEEYSTMVASVY LGILSAVTLFVFAKPEFTKELTAKLRFMKKE"	
gene	complement(2573. 3442)	
CDS	/gene="PF0310"	
	complement(2573. 3442)	
	/gene="PF0310"	
	/note="Function Code: 15.2 Other Categories: Drug and analog sensitivity"	
	/codon_start=1	
	/transl_table=11	
	/product="daunorubicin resistance ATP-binding protein homolog drrA"	
	/protein_id="AAL80434.1"	
	/db_xref="GI:18892260"	
	/translation="MGERVDFSCGWRNRVWRILIKDRSSSTFWSYANNSWCHGEGK MKAIVVEDLEKDYGVKALKGISFEVKEGEIFGLGPNAGKSTTLKIATLTLIPVGG RAIEFGYDVAEEAEYKRLISYLEPAGAYKNDGYEYLEPAKIIYAKGKYYKKLE TGEVLSGIEERLKDRISTYSGKMTKRLIARALMPEKAIIDIPRASGDIINASIR OTIKQPARRKGTPTLLSSNMLEVEPLCDRYALLINKRIVEIGTPRELKYNAMENLE EYFMSVGGKAYE"	
gene	complement(3307. 3657)	
	/gene="PF0311"	
CDS	complement(3307. 3657)	
	/gene="PF0311"	
	/note="Function Code: 16.1 Conserved Hypothetical"	
	/codon_start=1	
	/transl_table=11	
	/product="hypothetical protein"	
	/protein_id="AAL80435.1"	
	/db_xref="GI:18892261"	
	/translation="MMKIKLIGIITLTGFTLVIRPPDYRACMTLPDGEYCSGI NMPELLETGVIVYAGTLITLGNFRKCKERKGFILAGGLGIFGIGYSRIEVALFP GAMLVTIGVMEVRK"	
gene	complement(3724. 5091)	
CDS	/gene="PF0312"	
	complement(3724. 5091)	
	/gene="PF0312"	
	/note="Function Code: 6.7 Energy Metabolism: Glycolysis"	
	/codon_start=1	
	/transl_table=11	
	/product="Adp-dependent glucokinase"	
	/protein_id="AAL80436.1"	
	/db_xref="GI:18892262"	
	/translation="MPWEEYKRNATKATKSVKYGVGLTYNTNDAIKYDSHL EERLIKREYIKYSRELDKINTYSQIGSTIMSIRRKAAELFVESCVPRTMKR WGNELRMGQAGIMANLIGGVGVPIVAVPOLSRLOANLFDGPIYPTLEGEYK LHPKERSGDEENCIIYIFEPGFRFEPEARENREIISADYNTTLTIRREFRS FSEYIKNVQALIIISGLQALKEKYEFEKSNLEVLNREIIPVHLEFAFTDEKVR EELTNLVGMFYSVGLNEVALKSEIIEIGERKLAKELAHDPVPIAVTEAMLAKT GVKRIHHTGYLALIEYGEVRODLFPAALAAKAKKNTISFEIRATNSVY NERATQVEELRLREYKGEIGVEGQALFPTKIYAKRSTVGIGDTLSSAFIGE FSTPL"	
snRNA	complement(5145. 5199)	
	/note="small nucleolar RNA, posttranscriptional modification of rRNA"	
gene	complement(5192. 5449)	
CDS	/gene="PF0313"	
	complement(5192. 5449)	
	/gene="PF0313"	
	/note="Function Code: 4.6 Cellular Processes: Protein and peptide secretion; (sec11)"	
	/codon_start=1	
	/transl_table=11	
	/product="signal sequence peptidase I"	
	/protein_id="AAL80437.1"	
	/db_xref="GI:18892263"	
	/translation="MYENPLKYKPIIIRVREIKTISIEGREGICFTWGDNNPVPDY PPYGMDCVPGAIKAKALVVPRIIGISIKVREFLGIG"	
gene	complement(5504. 5713)	
CDS	/gene="PF0314"	
	complement(5504. 5713)	
	/gene="PF0314"	
	/note="Function Code: 16.1 Conserved Hypothetical"	
	/codon_start=1	
	/transl_table=11	
	/product="hypothetical protein"	
	/protein_id="AAL80438.1"	
	/db_xref="GI:18892264"	
	/translation="MHGNMNSREDKNINGGVYKERAVERIISTVLVLTFSIYYGM RIALHTDPLVAVVIOLVHETCLLOR"	
gene	complement(5641. 6186)	
CDS	/gene="PF0315"	
	complement(5641. 6186)	
	/gene="PF0315"	
	/note="Function Code: 16.1 Conserved Hypothetical"	
	/codon_start=1	
	/transl_table=11	
	/product="hypothetical protein"	
	/protein_id="AAL80439.1"	
	/db_xref="GI:18892265"	
	/translation="MTFGLVNTYDTRKLHEAHIRAIARAAVAYAEHRLALIGFPF DSDDEVEAVERASTTIGEGKRYLELAKSKNFILDPKRGPPPOFCEVATATSKP SKEKIVSRVAVEMALRGSKMLIIGLGHGHPKPEFEISKXVHLDTDOGREISTLETCT AGITPARIKITMEALMKKKP"	
gene	6254. 7306	


```

CDS
      /gene="PF0316"
      /locus="6254..7306"
      /note="Function Code: 16.1 Conserved Hypothetical"
      /codon_start=1
      /transl_table=11
      /product="hypothetical protein"
      /protein_id="AAL80440.1"
      /db_xref="GI:18892267"
      /translation="MKDIVESEVELVRGMIEAKIVENIQENEDKVALCTAEI
      TRDSELEDAIYELRKIDHDKNAIGYEVAKYALDGYEESLEFEBAVKIKMLN
      PYDKATLEIARLALAGFVDSLKLEPNAPAEIIEIMLEIKTDLIOLAEIEA
      SGDAIDSENAKFEYMAODIYKIKIGOSEGIKEIKELTRLMHTSYDIRIAVEG
      RFOALHEIEKPRGNNVPILETTPMAKTYFEELKIKETIMELTSKISDEKPF
      KYVQLLEIDDLTKAYEIAITLESQELREKALVLTGLKMIIDSEVAFVEKIFPLSD
      ERKEEFINFLKNIPIA"
      /gene="PF0317"
      /locus="7303..8145"
      /note="Function Code: 16.1 Conserved Hypothetical"
      /codon_start=1
      /transl_table=11
      /product="hypothetical protein"
      /protein_id="AAL80441.1"
      /db_xref="GI:18892267"

Query Match      44.28; Score 825.2; DB 1; Length 13354;
Best Local Similarity 66.68; Pred. No. 151;
Matches 1245; Conservative 0; Mismatches 608; Indels 15; Gaps 4;

1  ATGACGGCGATCGAATGGAACACGAGACCTTTTCTAAGTCCCTACCTGGCGACCGG 60
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  8179  ATGAGTTCATCGAATGGAACGAAAAACGTTTGCAAGTTTGCATACCTGAGTATCCC 8238

QY  61  AGGATACGGGGAACTTAATCGCTACACCTGACGAGGCGCAACATGAGACCAACAG 120
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  8239  AGCAACAAAGGAGAAATGTGTGCTATGTTAAACAAAGCCCAACTAAAGGACAAACAA 8298

QY  121  TAGGAGACGCGGTGTGTGTAAGACCTGAAACGGGCTCAAGGGCTTCATCGAGAAC 180
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  8299  TATGAAAACCAATTTGTTATGGAACCTCAAGAACATCAAGAAAGGTTTATGAAAT 8358

QY  181  GCCTCAATGCCGAGATTTGCCAGACGCGCAAGGCTCCCTTCACTGCTTTAAAGAG 240
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  8359  GCCACATGCCAAGAAATTTCCCGGANTGTAAAAAGTACCTTTTATGAGGGCTATAGAA 8418

QY  241  GAGAGAGAGAGACCGAGATATGGTGGCCGATATCCAGACCTGACGCCCAAGAAATC 300
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  8419  GAAAAAGAGGTAGATGATATGGTGGCAGATTTAGAACTTTAAGTACCAAGAAATC 8478

QY  301  CTCTCACTAAAAACGCGCTCGATGAGATGGAAGAGATTCAGAGGACTCTTAAT 360
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  8479  CTGGAAGCTTAAGACATTAAGATCTTGGAAATGGAAGACGAACTCAAGAAATCTTCAT 8538

QY  361  GTGCGCTTCAAGAGAGAGAGATGAGAGACTTCGCTTTTGACAGACGTCGCCGCTGG 420
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  8539  GTTGGCTTTAAGAGAGAGAGATGAGAACTTTATTTGAGAGATATGTTCCAGGCTGG 8598

QY  421  TTGACCAATATGGGATCTTGTATGAGAGAGACGAGCTTGTGGGTTCTTGACACTGAG 480
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  8599  TTTGATGATCTTGGATCTTGTATGAGAGAGACCAACATTCGAGTTTGTGACAGGAG 8658

QY  481  GCGGAGAGATATTCGAGATTCGAGAGAGCGGAGGTTTCAAGTGGGCTCTGGACGGC 540
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  8659  AGTGAAGAGATTTAGAGAGGTTGGAAGAGCGCGGCTTTCTTCAGAGAAATTTGGCAAGA 8718

QY  541  GATGCAATGTTGTGAAGCTCCGACCGGAGAGGAGACCAAGCTCCGCTGTTCAAGTTTC 600
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  8719  GACAAATATGTTGTGAATGTTCCCAATAGAGA---GATATCCCAACAAATCTTTAAGTTC 8775

QY  601  TACGACATAGTCTTATGGAAGAGCGGAGAGAGAGAGAGCTTTCAGAGGCTCTCCCTTC 660
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

```

Db  8776  TGGGACATCTATATCTGGGAGACGGAAGAAAGAAAGATGTTGAAAGATATCTGTTCC 8835

QY  661  GAGGCGGTTGACTCCGACGGAAGAGAAATACTCTGAGGGCGCAAGAAAAAGCGGTTCC 720
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  8836  TACGAGTGTGATTCGATCGAGAAAGATTTTGTCTATGAAAGCCCGAGAAAGATAC 8895

QY  721  ATCAGGACACGACTGCTGTAACCTCT-----GGAGCGGAGACTTAAACCATCTAC 774
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  8896  ATGAGCGACACCAATTAAGCTTTACATCTGATGGAAGAAAGATGAGGACTTCTGTAT 8955

QY  775  GAGGCGCCCTCGACTCTGGGAGACCAAGCTCACGGAAGAAAGAGTCTCTCTCACT 834
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  8956  GAAGTTGACAGAGAGATTTGGCCAGCAAGATTAAGAAAGTGAAGTACTTTACACTC 9015

QY  835  CCAGATCGCGCAGAGGTTAACTCTGCTGCTGGAGCGGAGCGGAGCGGTGTTTAC 894
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  9016  TTTGAAGAGAGAGAGCGCTTAACCTATTCATTTGGAGATGGGAAATTAACCCATAGCAAG 9075

QY  895  GGGGACCACTGGATTTACGGGGCTTGACGTACAGGANTGGCAAAAGCATTCCTCCATCAT 954
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  9076  GGAAGCATTTGATTAATGAGATTTAGCT---TGATGAGATGCTGTTTACCTTAAAGAG 9132

QY  955  ACCGCGACGAGATAGCGAGCTCTACCTCTACGAGCGGACCTGAAACAGGTCAACGGA 1014
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  9133  ACGGCGACTAATTAAGGAGGCTTTACCTGGAGCGGTGAGAGAAACACTCACTGAC 9192

QY  1015  TACACGGGCGCATATTCAGGAAGCTCAAGACCTTGAGCCGACGACACTCCGCTTCAAG 1074
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  9193  TATTAAGATCCAACTCTCTTAAGCAAACTTTCGAAACCAAGTCACTCTGCTTATTA 9252

QY  1075  AGCAAAAGACCTCGAGATTAAGCGGCTGTACCTCAGCGCGGAGGTTAAAGG---GAGAG 1131
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  9253  AGCTGATGATTAAGATTAATGATGTTGTTATGAAACCCGACACTCAAGAGGAGGAGAAA 9312

QY  1132  GCCCGGATATGATCTTCTGTCACAGCGCGGCGGAGAGGATAGGACACCGCTTGCCT 1191
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  9313  GCCCGGATATGATCTTCTGTCACAGCGCGGCGGAGAGGATAGGACACCGCTTGCCTCAAG 9372

QY  1192  TACGATGATGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1251
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  9373  TATGAATGCAATTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9432

QY  1252  AGCGAGGCTATTAAGCAAGACTTCGCGCTCCGCTCTGGAAGAGAGCTGCTGAGAGAC 1311
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  9433  AGCAATGATATGATGAGAGCTTTCGACTAAGATTAACCAAGACTGAGCTGAGAGAG 9492

QY  1312  TTTGAGACATTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1371
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  9493  TTCCAAGATATTAATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9552

QY  1372  GAGCGGCTTGAATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1431
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  9553  GAGAGATTTGATTCAGCGGAGATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9612

QY  1432  CAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1491
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  9613  CAAAGTGAATCTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9672

QY  1492  TACGCTTCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1551
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  9673  TATGCAATTTTCGAGCAATTTGCTTGTGATTAAGGAGGAGGAGGAGGAGGAGGAGGAG 9732

QY  1552  GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1611
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  9733  GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9792

QY  1612  ATACTCTAATCACTGCTGTTGAGAGTACCGCTGTCCGCTGACCAAGGCTTATGTTTC 1671
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  9793  ATTATTAATTAATCAAGTCTAGAGAGTATTCGCTGTCCCTGTGATCAAAAGTGTATG 9852

QY  1672  TACAGGCTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1731
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  9853  TATCAGGCTCTTAAGAGATCTAGAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9912
```


	Matches	1086;	Conservative	0;	Mismatches	770;	Indels	32;	Gaps	7;
QY	1	ATGACGGCATCGAATGAAACACGAGACCTTTTCTAATGTCGCTTACCTGCGGACCCG	60							
Db	9370	ATGGACAAATTAATATGGGACGAGAGTACTTTTACGAATTTGGCTTCTTAAAGACCCG	9311							
QY	61	AGAT-----ACGGGMAACTTAATGCGGTACACCTGACGAGAGGCCACATGAAAGAC	114							
Db	9310	AGATTTTCAGAGGATGGATTAAATATAGCTTATGTCTCTACAAAGGCGAGTTTGAAGAT	9251							
QY	115	AACAAGTACGAGACGACGCGTTGTTGTTGAAGACCTTGAACGCGCTCAGGCGCTTCATC	174							
Db	9250	AATGAATAACAACATACCATTGTGTGAGAA-----GAAAAGGAGACGAGAGTTTATYT	9197							
QY	175	GAGAACCCCTCAATCCGAGGATTTGCGACAGCGGCAAGACGTCGCGCTTACGTCGTT	234							
Db	9196	GAAATGCGACGATGCGCGCTTTCTCTACGSGGAGGAAATGATTTTGTAAAGAG	9137							
QY	235	AACGAGAGAAAGAGGACCGAGATATATGGTGGCCGATATCCAGACCTGAGCGCCAG	294							
Db	9136	CATAGGAGTAAGATGTTTAACGAGATATTTTATAGCGATTTAAATTCATGATGAAAGG	9077							
QY	295	AAATGCTCTCACTAAAAAGTCGCCCTCGATGCAATGGAACGACGATTTCAAGAGACTC	354							
Db	9076	AAGATTTTGGATTTAAAGGACATTTTGGATTTAACGTGGATTTAAAGATGACAGAAATTT	9017							
QY	355	TTACTTTGCGCTTCAAGAGAGGAGACATGAGACCTCGATTTTGAACGACGACGTCGC	414							
Db	9016	TTGATTAACAGGGAATTTAAAAATATGAAAGTAAAGACCTCATTTTGAAGATGAGGTTCT	8957							
QY	415	GTCGTGTTTCACAATATGGAATCTCTTGATGAGAGAAAGACGACGCTTGGGTTCTTAC	474							
Db	8956	GTCGTGTTGATGATTAAGAGATTTTGGATTTTAAAGGATTTAAAGATGACAGAAATTT	8897							
QY	475	ACTGAGGCGCGAGAGATATATCGACAGATTCAGAAAGCCGAGTTTTCAGTGGCGCTCGG	534							
Db	8896	GTGGAAGTGGAGATATCTTCTGATGATGTTTCTCGAAGAAATTTTCTACGCTATCGG	8837							
QY	535	CACGGCGATCGATAGTTGTGACGTCCCGACCGCAGGGGGAGCACGCTTGCCCTGTTCC	594							
Db	8836	CATGAGATGATGATTAATCTATTAATGTTCTTAAAAAGGTGGACAATGTTCCCAAGTTT	8777							
QY	595	AAGTTCTACGACATGATGCTATGGAAGAGCGGGGAGAAAGAACTTTGAGAGGGCTC	654							
Db	8776	GGACTTTTGGCATATATACGTTTACGAGAGGAAAAAGAGAAACTGTTTGAAGATGTT	8717							
QY	655	TCCTTCGAGCGGTTGACTCCGACGAGAAAGAAATCTCTGAGGGCGAAGAAAAAAG	714							
Db	8716	TCGATATCTGCTATGATTTTAAAGGAGATGGTGCTACTTTACGGTAAACCCAGGAAA	8657							
QY	715	CGGTTTCATCAGCGACGACGACTGGCTGTACTCTGTGGACG-----CGAGCTTAA	764							
Db	8656	GAAAAAGCACCGAATGATTTTCACTATCTGTGGGACGGGAGAAAGTTATCCCTTGG	8597							
QY	765	ACCGATTCACAGGGCCCGCTCGACAGTGTGGGAAGCCAAAGCTCAGAGAAAGGTCTA	824							
Db	8596	ACAGAAACTCTGTGTATACATTAACGACGAGGGGAAGTATGATGAAA--AGGAAATGTTA	8538							
QY	825	CTTCCCTCACTCAAGATCGCGGCGAGGTTAAACCTCTGGCTCTGGGACGAGGAAAGCCGACG	884							
Db	8537	CTTTACTACTGTGCAAAAAGAAAGGAGATTAAGCTTTTATAGCTGAGGAAAGATTTGGC	8478							
QY	885	TGTTGTTAACGGGCGACAC--TGATTTACGGGCTTACGTCACG---GATGGCAAGC	938							
Db	8477	TTCTATTTCAGAACTTAATCATATGGGTAGGAAATTTTACCTGTACGGAAGCGGAAAGT	8418							
QY	939	ATTGCTCTCATCATGACCGCCACAGAGATAGGCGAGCTTACCTTACGAGCGGAGCT	998							
Db	8417	GGTTTTCACATAAGAGACTGATATTTATCCAAAGCAACTCATGTATGGCATGAAGGCT	8358							
QY	999	GAACAGGTCAACCAATTAACAGGGGCGGATTTTCAGAAAGTCAAAGACCTTCGAGCGAG	1058							
Db	8357	TGACACGTGACAGATTAACATAAGCTATGATGATGATGAGCTTGATGTGAAGCATATTA	8298							

QY 1059 GCATTCCTCCGCTTCAAGACCAAGACCTCGATAGACGGCTGTACTCTCAGCCCGAGGT 1118
 DB 8297 CAATTTTAAATACAAAGCTCTGATTGTAATTTGATGATGATGATGATGATGATGATGAT 8238
 QY 1119 TAAAGAGG---AGGCCCCGCTGATAGTCTCTCCACGGCCGCGGAGGATGTA 1175
 DB 8237 AGAAGAGGCAAAAGGCTCCGGCTTTGGTTTGTTCATAGCGGCTCCAAAGGGGATGTA 8178
 QY 1176 CGACACCGCTTCCTGCTACGAGATGACCTGATGCGGAGCAAGGAGCTCTACTGCTGCT 1235
 DB 8177 CGGGATATCTTTAAATACGAATTTCAACTTATGCGCATAGAGATTTTATGTGCTATT 8118
 QY 1236 CGTGAACCGCGCGCGGAGCGAGCGGCTATAGCGGAAGACTTCGCGCTCCGCTCTGAGAG 1295
 DB 8117 TGTAAATCTTAGAGGAGATTAATGCTACAGCGCAAAATTTGCAATGAGTGTGCTGAAG 8058
 QY 1296 GACTGCTTGGAGACTTGTGAGACATTAATGAAAGGCAATGAGAGTCTCTCAAGCTGTA 1355
 DB 8057 AACAGGCTAGAAAGTTTTCAGACATTTAAATGGGTAGAAAATTTCTTGAATTTAGA 7998
 QY 1356 ACCGAGGCGGAGGAGGAGCGCGTGAATTAACGGGCAATAGCTACGGCGCTTCATGAC 1415
 DB 7997 ACCAAACGCTGATAGAGAAAGATTGGAATTAACGGGTATAGCTATGCTGGTTATGAC 7938
 QY 1416 CACTGGGCTTGTACTCAGACGACCTTTCAAGGCAAGATTAAGCGGAAGGCAATAG 1475
 DB 7937 TAAATGGCTTGTACTCAGACGACCTTTTAAAGCTGGCAATAGTGAATTTGCAATAG 7878
 QY 1476 CTACTGGCTCACACGCTACGCTCTCTCGACATAGGGCTGTGATGACGCTCGAGTAT 1535
 DB 7877 TTACTGGCTTACAGATTTTTCAGATTTTTCGACATGATTTATGTTGACAGAGATGAT 7818
 QY 1536 CGGCGCAATCCGTTAGAGAAAGAACTTCAAGAGCTCAGCCGCTGTTCTACGCTCA 1595
 DB 7817 TGGGGAATATCCACTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7758
 QY 1596 GAAGCTGAGAGGCGGAGTACCTCAATCCATCCCTTGAGAGCTACCGCTCTCGCTGTA 1655
 DB 7757 AAATGTGAGAGGCGGAGTACCTTCTTGAATTTCAATTTAGAACATCAAGATCTCCCTGTA 7698
 QY 1656 CCAGAGGCTTATGTTCTACACGCTCTCAAGAGATGAGGAGGAGGCTTACATAGCAT 1715
 DB 7697 CCAGAGGCTTATGTTCTACACGCTCTCAAGAGATGAGGAGGAGGCTTATAGCAT 7638
 QY 1716 ATTAAAGCGGCGCGGAGCGGAGCGGCTCCGCGAAGCGCGGAGCAAGCGCGAAGCG 1775
 DB 7637 CTTCAAAAAGAGAGCTCAGCGGAGCGGCTGATGAACTCTGCTCAGAGGCGAAGAAAG 7578
 QY 1776 CTACAGGCTCTTCAATAGTTCTCGAGCGCAAGCTCAAGAGATGAGAGGAGGCTTGA 1835
 DB 7577 GTAAATCTATTTTATGAAATTTTGAAGAAAGCACTCAAGAGATGAGAGGCTTGA 7518
 QY 1836 GGTAGAGAGATTAATCAAGGAGGATGGG 1863
 DB 7517 TGTGAGGAGGATTTTGAAGAGGAGAG 7490

TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Tanaka, T., Masuda, S., Funahashi, T., Tanaka, T., Kudoh, Y.,
 Yamazaki, J., Kushi, N., Oguchi, A., Aoki, K., Kubota, K.,
 Nakamura, Y., Nomura, N., Sako, Y. and Kikuchi, H.
 Complete genome sequence of an aerobic hyper-thermophilic
 crenarchaeon, Aeropyrum pernix K1
 DNA Res. 6 (2), 83-101 (1999)
 2 (bases 1 to 247695)
 Tanaka, T., Hino, Y., Kawarabayashi, Y. and Kikuchi, H.
 Direct Submission
 Submitted (14-DEC-1998) Yutaka Kawarabayashi, National Institute of
 Technology and Evaluation, Biotechnology Center, Nishihara 2-49-10,
 Shibuya-ku, Tokyo 151-0066, Japan (E-mail: kyutaka@kazusa.or.jp,
 Tel: +81-3-3481-8951, Fax: +81-3-3481-8424)
 Kawarabayashi, Y. is officially affiliated with the National
 Institute of Bioscience and Human Technology, Tsukuba, Ibaraki
 305-0046, Japan.
 Nomura, N. and Sako, Y. are at Kyoto University, Sakyo-ku, Kyoto
 606-8502, Japan
 The other authors are at the National Institute of Technology and
 Evaluation, 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan.
 All the sequence with length 100 codons or more between ATG or GTG
 and stop codon are defined as CDS
 Homology analysis is performed by Smith-Waterman algorithm against
 Genbank and Genpept release 109; EMBL release 56.0; SwissProt
 release 36.0; PIR-Protein release 57.0; and OMT release 31.0.
 E-mail address for comments and questions: genome@nitech.go.jp
 Restriction map, ORF organization, sequence alignment and more
 information are available at W.M.W. site of Biotechnology Center,
 URL: <http://www.mild.nitech.go.jp/>.
 Location/Qualifiers
 1..247695

FEATURES

Source

/organism="Aeropyrum pernix"
 /strain="K1"
 /db_xref="taxon:5636"
 complement(721..798)
 /gene="APECRNA43"
 complement(721..798)
 /gene="APECRNA43"
 /product="rRNA-His"
 /note="codon recognized: CAC; anticodon: GTC"
 /anticodon="(pos:761..763,aa:His)
 complement(980..1501)
 /gene="Ape2245"
 complement(980..1501)
 /gene="Ape2245"
 complement(980..1501)
 /transl_table=1
 /product="173aa long hypothetical protein"
 /protein_id="BA81257.1"
 /db_xref="GI:5105946"
 /translation="MLNDVSIIEVNDPEAGVPCVCHNIPMASPROPLATHNIRPSP
 SLGEVGYDLPPRRRAGGEGVCEVLELRETEGDSRORVNTPPRPIAAGQWMD
 VSNAAPEPEALOSTIPGPGVTPAPRRPPHNNRPHRDEIRPLAQPPRTVAQIDL
 PNOPTPTPIHNR"
 983..1549
 /gene="Ape2246"
 983..1549
 /note="motif-eukaryotic putative RNA-binding region RNP-1
 signature"
 /codon_start=1
 /transl_table=1
 /product="188aa long hypothetical protein"
 /protein_id="BA81258.1"
 /db_xref="GI:5105947"
 /translation="MYVDGSCWLVSEVYTTIGSGSLGELIDLVESGSRVVDVR
 RPPSRHRPWRASRDLEGLRGKVAITLIPALGGRKSRFDVDPISRALFSCESESG
 FNAIYALASPPAMRGLIVAIYLAEGAPVWCSERLPPWRKRVADMLVARGFR
 VLHILDRDSIVHRGTCAAMLGSPSSV"
 complement(1536..2120)
 /gene="Ape2247"
 complement(1536..2120)

RESULT 9
 AP000064/c 247695 bp DNA linear BCT 06-APR-2000
 LOCUS
 DEFINITION
 Aeropyrum pernix genomic DNA, section 7/7.
 ACCESSION
 AP000064 BA000002
 VERSION
 AP000064.1 GI:5105945
 KEYWORDS
 Aeropyrum pernix (strain:K1) DNA.
 SOURCE
 Aeropyrum pernix
 Archaea: Crenarchaeota: Thermoprotei: Desulfurococcales;
 Desulfurococcales; Aeropyrum.

REFERENCE
 AUTHORS
 1 (sites)
 Kwarabayashi, Y., Hino, Y., Horikawa, H., Yamazaki, S., Halkawa, Y.,
 Jin-uo, K., Takahashi, M., Sekine, M., Baba, S., Anai, A., Kosugi, H.,
 Hosoyama, A., Fukui, S., Nagai, Y., Nishijima, K., Nakazawa, H.,

gene
CDS

[illegible]

REFERENCE	ORGANISM	DEFINITION	LOCUS	RESULT 10
1228	Yersinia enterocolitica	10256 bp DNA	linear	BCT 22-NOV-1999
1287	Yersinia enterocolitica	10256 bp DNA	linear	BCT 22-NOV-1999
14668	Yersinia enterocolitica	10256 bp DNA	linear	BCT 22-NOV-1999
1288	Yersinia enterocolitica	10256 bp DNA	linear	BCT 22-NOV-1999
14611	Yersinia enterocolitica	10256 bp DNA	linear	BCT 22-NOV-1999
1348	Yersinia enterocolitica	10256 bp DNA	linear	BCT 22-NOV-1999
14551	Yersinia enterocolitica	10256 bp DNA	linear	BCT 22-NOV-1999
1408	Yersinia enterocolitica	10256 bp DNA	linear	BCT 22-NOV-1999
14194	Yersinia enterocolitica	10256 bp DNA	linear	BCT 22-NOV-1999
1468	Yersinia enterocolitica	10256 bp DNA	linear	BCT 22-NOV-1999
14134	Yersinia enterocolitica	10256 bp DNA	linear	BCT 22-NOV-1999
1528	Yersinia enterocolitica	10256 bp DNA	linear	BCT 22-NOV-1999
141374	Yersinia enterocolitica	10256 bp DNA	linear	BCT 22-NOV-1999
1585	Yersinia enterocolitica	10256 bp DNA	linear	BCT 22-NOV-1999
14314	Yersinia enterocolitica	10256 bp DNA	linear	BCT 22-NOV-1999
1645	Yersinia enterocolitica	10256 bp DNA	linear	BCT 22-NOV-1999
141254	Yersinia enterocolitica	10256 bp DNA	linear	BCT 22-NOV-1999
1705	Yersinia enterocolitica	10256 bp DNA	linear	BCT 22-NOV-1999
141194	Yersinia enterocolitica	10256 bp DNA	linear	BCT 22-NOV-1999
1765	Yersinia enterocolitica	10256 bp DNA	linear	BCT 22-NOV-1999
141134	Yersinia enterocolitica	10256 bp DNA	linear	BCT 22-NOV-1999

TITLE Direct Submission
JOURNAL Submitted (08-NOV-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

FEATURES location/Qualifiers
source 1..10256

gene /organism="Deinococcus radiodurans"
/strain="R1"
/db_xref="taxon:1299"
/chromosome="1"
100..1017
/gene="DR0164"
100..1017
/gene="DR0164"
/note="similar to GB:D14846 PID:286165 percent identity: 49.11; identified by sequence similarity; putative"

CDS /codon_start=1
/transl_table=11
/product="endo alpha-1,4 polygalactosaminidase, putative"
/protein_id="AAf09753.1"
/db_xref="GI:6457833"
/translation="MRPLFVLVPLSAVLASGSGTPSPAATVFPPAPQPTTLA AQAAPPRLPAGKLAWDMOIGATEKVALPGVSLLDIDGETSAKYADAKAGGV YKCYLNVGSEYRPDAAYPSIKIQPNMPEDFVIDIRFRSGYLAKGIDRR IALCAKGFPAVEPDNLQDNQNTSGVISKODQLDEGMILADRAHAGLIAQNGPD TVAQDGKRLMDLFEDGVINESCQRKECGPILEYRKRLALNVEYROADLNCAIM NSLVGNALRFDLSTVGKLOKAYKRVSCTG"

gene complement(1079..3046)
/gene="DR0165"
complement(1079..3046)
/gene="DR0165"
/note="similar to GB:AU009126 percent identity: 56.01; identified by sequence similarity; putative"

CDS /codon_start=1
/transl_table=11
/product="acyl-peptide hydrolase, putative"
/protein_id="AAf09754.1"
/db_xref="GI:6457834"
/translation="MNSETPAPCPBSLLAAPSPDPQPVDCQVAFLYAQISEEP AKPKDKPRPRTRNSGLMLSEGARPLTHAEFGSDAPRMSPDGONLARVRSAGEVK AALMLPLKGGEARSRVTHFNKGVSQPWSPDGRITATTADTEKDKEGEARLYVR PYVRANQADMILPERPALMLYDEADAILREMYAPEISIGALSMPDSRGVLYOSEBE WOSORAMQODYDLPLPTADAAPAPOKILLMNSAANGIAPAPDGOERFLIGRPAGKGT EAHVILIENGQHRRLDTGDHPVDGAGDDCHYGARPEGPRMLDPTGLLESTSVRS VGLFTAITGSGYKAYDHDPOGVISAIFANHEGVALIHESATPREVEYLKQRTDILA RFFPPAREPQRTVEETELGEGEAVLLPEEGOKVPALLNHGPHTDYGGTFHERQL MAWGSQVCYNPBGVSXGYGAWDALYGRMGVTADADLLNFEDRCLEAVRIDAAQT AAAGXSGEMTYNIHTGTRYFOADAIIDRCLISNISGSGSIDGLRFMWDELIDFSR RAALKLTMDLSPLYOVENVKTPLIIVSYVDHRCRPVGAQOWYAAALHKHQVPYRFVF PENNHELRSRGRDRRLTRIIFYEAMLERLL"

gene 3098..3352
/gene="DR0166"
3098..3352
/gene="DR0166"
/note="similar to SP:P31787 PID:1323021 PID:1620454 PID:1170970 percent identity: 66.09; identified by sequence similarity; putative"

CDS /codon_start=1
/transl_table=11
/product="acyl-CoA-binding protein"
/protein_id="AAf09755.1"
/db_xref="GI:6457835"
/translation="MNEEQAOEVNLTSKRGNDVILKLIALYKQSAGDVDKRPGC EDVVGAKRYDMAKLGKTQEQDAQOEIVALLYELKAKDQ"

gene 3663..4649
/gene="DR0167"
3663..4649
/gene="DR0167"
/note="identified by Glimmer2; putative"

CDS /codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAf09762.1"
/db_xref="GI:6457842"


```
gene
/translation="MPSANVSPPCSGVGGMGKRAAEASKPEHPQIPVKLPPTAP
DALAAKAKMRDLAAIYVAALPGRTHSLMAVPPVDLKEPFLGWRDGAPEHNVIL
INSAPPERQRTLAHEIGHAILLDDDLSDIDYAGEERLEQVETICNVAAAIL
MPEVIAELERFEGTRALAEIARAEVASASSALYALTEQTPVYIAYACAPGKPR
BOASDEDAGPTEKVKLVIRASSTRGVKTLAGSTPVADHPALALATGCEVRES
YVPSRGRKAEVDAYPSRGIYVNSFEPDARLRKDSQADRDEPDAAQ"
4646..5515
/gene="DR0168"
CDS
/feature="DR0168"
4646..5515
/note="similar to GB:L06494 SP:P26282 GB:U01376 GB:X68776
GB:X68777 percent identity: 62.16; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="dihydropteroate synthase"
/protein_id="AAF09756.1"
/db_xref="GI:6457836"
/translation="MNTGRIHELTFFGPPVSAEKGPNQTVYRPGCYMGILNTPD
SPSDGQHLQDLDAIATARRNDVTFIDIGESTRGAEPVDATIEDRYLPIIRA
LGEHVLLSVDTMRPEVAEALRAGAHVNDVTGLRDPAMIRVCADAGAAACVMHOG
EPRTQHQPVYSDVVAEYHGFLERAAEVKSQVPSVLLDPGIGKTLIDHNLILRA
VDLDSGDPVLIASRKKITDILAGVEADRDPSIALHLHAARCGAALVRLHAAG
EMVQALRYEALAA"
5540..5896
/gene="DR0169"
CDS
5540..5896
/note="DR0169"
/note="similar to GB:D26185 SP:P28823 GB:M34053 PID:142411
PID:467467 percent identity: 62.18; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="dihydroneopterin aldolase"
/protein_id="AAF09757.1"
/db_xref="GI:6457837"
/translation="MSRYVLEGLFEFHAHHGVPAAGVLCARVVDATLHSPAGIAD
LGOAVNYAAVAIRREYTEKTHQLEVLADRIADRLLSDPSRLERVYRHKPRAPL
PGIFRDVYAEIEKERT"
5893..6408
/gene="DR0170"
CDS
5893..6408
/note="DR0170"
/note="similar to GB:D26562 SP:P26281 GB:L06495 PID:146013
PID:473801 percent identity: 56.86; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="2-amino-4-hydroxy-6-
hydroxymethylidihydropteridine pyrophosphokinase"
/protein_id="AAF09758.1"
/db_xref="GI:6457838"
/translation="HTRAFIALGANLGPQATLRALTELGLAGEVRSALYRTAPV
GGPGQPYLNAVALLELALSNAPLLAALHALEADAGTRDVRWARVLDLILYGE
DVSDPALTLPHPRAMERGFYLAFLDAPLSIAHPRTEMVAAQALRRVLAGVEVSEL
SLCMEIISGSE"
complement(6362..7054)
/gene="DR0171"
CDS
complement(6362..7054)
/gene="DR0171"
/note="identified by glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF09763.1"
/db_xref="GI:6457843"
/translation="NNLEGVHPLRLGLWCMEQTSRSSQGVDTTEVASALLDTRSR
LVLSAFLPTSLSAARDLGRPLNTFYQWOKLRLGLKMYCDGRPPLOYQEGSFE
IYELTLPAMHPODMQLHTENNRMDALSTADEABRHWGMVMOYRBOENALILE
HAASAGNTEPPQAPAVLDLGLSLDFEQAKAQRLDQLLAQYRRKPGSPHLIR
LTLTPASTEKAH"
7034..7849
/gene="DR0172"
gene
```

```
CDS
7034..7849
/gene="DR0172"
/note="similar to SP:050665 PID:1449319 GB:AL123456
percent identity: 49.77; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF09759.1"
/db_xref="GI:6457839"
/translation="MDTLKIGFSNGIOVKSKFLRPEPNLGLSIRSLDROXSAPL
PWCWLIHPSGLIYVPGILCTASNSFSRTORHYTEEDRLSRMORAPPAV
RVLIIFHLIDHIDRGMDFPEAHILVSRKEDRLSPARLMTGSATPKSVPAAP
LPLAGPRESADVMGDRVLLPTPGHPLSVLVKQPHGLDILAGDITVSEODLR
TGRIGIYVSPHIRKSMROYROLAQLRPLIYLPSPHSDSLRLQLRQNPFSLT"
complement(7904..8999)
/gene="DR0173"
gene
/note="this region contains an authentic frame shift and
is not the result of a sequencing artifact; similar to
SP:P28367 PID:216335 PID:231287 GB:AL009126 percent
identity: 66.01; identified by sequence similarity;
putative"
9254..9778
/gene="DR0174"
CDS
9254..9778
Query Match 8.9%; Score 166; DB 1; Length 10256;
Best Local Similarity 56.1%; Pred. No. 1.7e-22;
Matches 360; Conservative 0; Mismatches 270; Indels 12; Gaps 2;
1090 CGAGCGGAGGCACTCCGCTTCAGAGCAAGAAAGCTCGAGATAGCGGCTGTACCTCAG 1109
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1850 CGAGCGGAGGCGCGCTGACGTTTGAGACGAACTCGGAGGGGGGGCTGGCTGCT 1791
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1110 GCCGAGGTTTAAAGAGGAGGAGCCCGGCGTGTACTTCTTCACAGCGGCGCGGAAAGG 1169
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1790 GCCGGAAGT---GAGCAGAAAGTCCCGCGCTGCTCAATATACAGCGGCGCGGACAC 1734
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1170 CATGTACGACACCGCTTCGTACGAGATGACGTATGCGAGCAGAGGCTACTACTG 1229
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1733 CGACTACGCGACGCGCTTACCCAGAGTCCAGCTGATGGCGGCGCGCTACCGGCT 1674
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1230 CTGCTCGTGAACCGCGCGGCGGAGCGGCTATAGCAAGACTTCGCGCTCCGCGTCT 1289
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1673 GTGCTACTCCAAACCCCGCGCTCGTGGCTGCTACGCGGCGGTGGGAGCGGCATCTA 1614
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1290 GGAGAGGACTGCTTGAGGACTTTGAGACATTAATGACGCAATCGATGAGAGTCTCAA 1349
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1613 CGGCGGCTGGGACAGGTGGACGCGAGACCTGCTGAATCTTCGACCGCTGCTGGA 1554
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1350 GCTCGACCGGACGCGGACAGGAGCGGCTTGGAATTAACGGGCAATACGCTACGGCGCT 1409
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1553 AGCGGTGCGCGCTTGAGCGCGCAAAAACGCGCTCATGGGCGGCTACGCGGCTT 1494
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1410 CATGACCAACTGGCGCTTGACTCAGAGGACCTTCTCAAGCAGGAGAAATAGCGAGACG 1469
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1499 CATGACCAACTGGATCAACGGGACACACACCGCTTTAGGGCGCATCACGACCGCTG 1434
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1470 CATTAAGTACTGGCTACACGACTACGCTTCTCGAGCATAGGCTCTGCTACGAGCTGA 1529
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1433 CATCAGCAACCTGATTTGTTGGGGGACACGCGCATCAGCGGCTGCTTCTGAGACGA 1374
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1530 GGTCAATGGGGCCCAAT-----CCGTTAGAGAAAGCAACTCAGAGAGCTCAGGCC 1580
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1373 CGAGCTGGGGCTGGAAGCTTCTCGCGCGGGGCGGAGCGCTGGAAGCTTGGACCTTTCGC 1314
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1581 GCTGTTTACGCTCAGAACGTGAAAGCGCCGCTACTCTCAATCCATCGCTTGGAGACTA 1640
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1313 ACTGCAATACGTGGAAGAACGTGAAAGCGCCGCTGATTCGATTCGATTCGCTTGAACA 1254
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1641 CCGCTGTCGCTCGACCAAGAGCTTATGTTCTTACAGCTGCT 1682
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1253 CCGCTGCCCTTCGAGCAAGCGGACAGTGTGACCGCGCT 1212
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



```
/note="PAB1335"
/codon_start=1
/transl_table=11
/product="CHEMOTAXIS PROTEIN CHED"
/protein_id="CAB50455.1"
/db_xref="GI:5458969"
/transl_table="EMBL"
/LHALLPEARIRGHRGHPAKYVDTGLOLLKEVTKGASFRLEAKIFGAGMONTKS
DELKIGERNVOTAKRRLKRLGIRLVAEDTGGGRTTYLDTGVRMRKVIIGGOVIE
KY"
complement(10179..10817)
/note="PAB1334"
/codon_start=1
/transl_table=11
/product="CHEA HISTIDINE KINASE"
/protein_id="CAB50456.1"
/db_xref="GI:5458970"
/transl_table="EMBL"
/SAEFLKILAEKGVSKGTFVMDTEGLTILLOPKHSALINSLAVLMGMDGSMEE
LDMKRSALMEVGNILISAVTDLISLIGEPYLSPPRAESLYDEKRLPDLRNV
NSIVYFKSKFYREDICVESYFYIVPPESFTLVKLEKQISEVEHANEEN"
complement(10814..11452)
/note="PAB1333"
/codon_start=1
/transl_table=11
/product="CHEA HISTIDINE KINASE related protein"
/protein_id="CAB50457.1"
/db_xref="GI:5458971"
/transl_table="EMBL"
/MDIVANTIEVYKNDIFAKSALLETENIGASHAATASLSEMGGR
EVNISVPLDKVYKINRPEIVVEDVAVVYIELGDFSSHAFFADYDLRNFDM
GNPPTTLELDMVKSFMENGNILISAPANALSOPLGITIGTTPALTIIDFALID
LALADIGKFCDDTYIILQTDIKISGVFEFEEHFLPKPEMDKRLKELGLA"
complement(11436..13754)
/note="PAB1332"
/codon_start=1
/transl_table=11
/product="chemotaxis histidine kinase (chea)"
/protein_id="CAB50458.1"
/db_xref="GI:5458972"
/transl_table="EMBL"
/KABLINOIFRDHATLKGTATMGFKLSTFAHKMELPAIRKSOEELPEYVDVL
FLDALIEMVNNIEESGSDIDVSLFRADFLQKGAEKPEEKREVEAEKE
EKPERVEEYKGVYHVKYFQKDAFLKGVRAFLISDEKEGVNWTNPGRDVIEG
NLDKVIKRVSSDLPELEKIVYTHPEVEKEIIESEEEKGPEGEKEXILIRYF
SPDAPLKGRSLILODLESGIDIVOTNRGDIENGNIIEGEEGPEVYRTEDSKSI
VOLIKKHDPVENYVTEAEKKEKKEKBAKRPQPKOKRPYIEHPYKISII
KIDVGHIDLMNVGELVITKGRLEOIGRLGORELFEKLSLISLALVHLRNV
LPIAEVENKFPYVADLARKKKEVEFMEGADIEVDITLEKLDALVHLRNV
HGLEPERVYKLGKPRVYDELAKRKNHVVYIYVDDRGIDPEVKKALIERGLIT
PEASISLDEEALINILFLPGFSTAEKVTDSGRVGMVYKVMGNSISVQEVG
KGTTFILKPISMALIOALLIRVMDETVAIPINNILETIEVDPSTIKTGGKPVYLR
GLIIPVYMLHELGLPTPOVERPPATIVYHGAOKVAIGVDELHHRDIYIKSLKELS
NVRGPAATITLDSGVYLIIDIGLIGGYSG"
complement(13717..14823)
/note="PAB1331"
/codon_start=1
/transl_table=11
/product="protein-glutamate methyltransferase (chea)"
Query Match 8.18; Score 150.8; DB 1; Length 265118;
Best Local Similarity 50.88; Pred. No. 1.7e-19;
Matches 527; Conservative 0; Mismatches 487; Indels 24; Gaps 6;
```

```
58882 GAGAGCTTTCCGTTGGTATACATAGCATTTTACAGCTCAAGATCCGACAACTCCGACT 58823
970 GCGGAGCTCTACCTTACAGCGCGAGCTGAAGACAGGTACCGAATACAGCGGCGGATA 1029
58822 GAGCTTTACGCTCTCAGGATGGAAGAGCGATACACAGACTTTAAAGCCTGGA-- 58765
1030 TTCAGAACTCAAGACTTTCAGAGCGAGGACACTTCCGTTAG--AGCAAAACCTC 1086
58764 -TTAGGAGCTATTAACCTTCAAAACCGAGCACTTACAGGCTTACGCTAGCGTGAAG 58706
1087 GAGATAGAGGCGCTGATCTCAGCGGAGGTAAAG--GAGAAAGGCCCGGTG 1140
58705 GAAATAGATGCTCGATATGAGAGCGGCTGACTTTAGCTTGAAGGAGAGATATCTGGA 58646
1141 ATAGCTTCCACAGCGCGCGCGGAGGAGCATATGAGCAACCGCTTCTCTACAGATG 1200
58645 GTCCTGGAATTCACGCTGCGGCAAAACAGCCCTACGTTATGATGACGAGTTT 58586
1201 CAGCTGATGGCGAGCAAGGCGTACTCTCTCTGCTGTAACCGCGCGGCGAGCGAGC 1260
58585 CAGCTTTAGTTGCGCAAGGCTTCGTTGATCTTTTCAAAATCAGGGAAGTATGCT 58526
1261 TATAGCAAGACTTCCGCTCCGCTCGCTGAGAGAGACTGCTTGAGACTTTGAGAC 1320
58525 TATGGTGAAGAGTTCGCG--GACATAGGAGGTCATACGAGCGGATATCAAGAC 58469
1321 ATATGTAAGGCAATCAGAGATCTTCAAGCTCGAACCCGAGCGGAGAGCGCTT 1380
58468 CTATGAGAGTATGTTACAGAGCTTAAAGCCTTACTTATATGATGAGAGAGCTT 58409
1381 GGAATATACGGGCAATAGCTACGCGGCTTCAAGACCACTGCGCTTCACTAGAGCAG 1440
58408 GCGGTATCCGAGAGATCATACGAGGCTCATGACAAATGATGTTGGGACACAAAC 58349
1441 CTCTTCAAGCAGGAATTAAGCGAAGCGCATAGCTACTGCTACCAAGCTTACGCTTC 1500
58348 AGATTCAAAGCTGCGCTTACTACGAGCGCTCAATATCAACTGATGATGCTTTGGAACG 58289
1501 TCGGACATATGGGCTGCTGATACGAGCTGCGAGGTCATCGGCGCAATCCGTTAAGAAC 1557
58288 ACCGATATGCGCTTACTTACTTCTGCTCCAGATGAGATCGGCGATCTTGAGTATCCTA 58229
1558 GAGAACTTCAAGAACTCAGCCGCTGTCTACGCTCAGAACGTCGAAGCGCGGATATCTC 1617
58228 GAAGGCTATCGGGAGAAAGCGCATTTGAAGTACGCTCAAGCTGAGACCTCGCTCG 58169
1618 CTATTCACCTGCTTACGAGCTACCGCTTCCGCTGACCAAGCCTTATGTTTACAAAC 1677
58168 ATATTCACCTGCTCAGGAGACTACCGCTTGGCTTCCGAGGCTTTCATTTTACATA 58109
1678 GTGCTCAAGAGACATGGGCAAGGAGGCTCATAGGATATTCGAAGCGGCGCCGACGAGC 1737
58108 GCGTTCAAGATTTTGGGCAAAACCGTCGAGCTGCGGATATTTCCAGAGTACAGATATGAC 58049
1738 CACAGCGTCCGCGAAGCCGAGGACAGCGGAGCGGAGCGCTTACAGGCTTCTTACAGTTC 1797
58048 TTGATGATGAGTGGTGAAGCGCAAGCATAGGTTAAGAGCTGAGCTCATAGTGGGTTG 57989
1798 TTCGAGCGCAAGCTCAAG 1815
57988 TTTGAGAGGTGCTGGAAG 57971
```

RESULT 12
AX041922/c
LOCUS AX041922 265118 bp DNA linear PAT 23-NOV-2000
DEFINITION Sequence 817 from Patent W00065062.
ACCESSION AX041922
VERSION AX041922.1 GI:11340687
KEYWORDS
SOURCE Pyrococcus abyssi.
ORGANISM Pyrococcus abyssi
Archaea; Euryarchaeota; Thermococci; Thermococcales;

REFERENCE 1 (bases 1 to 265118)
 AUTHORS Forterre, P., Thierry, J.-C., Priet, D., Dietrich, J., Lecomte, O., Querret, J., Weissenbach, J., Saurin, W., Helling, R., Flament, D., Raftin, J.-P., Henneke, G., Gueguen, Y. and Rolland, J.L.
 TITLE Genome sequence and polypeptides of *Pyrococcus_1* (abissay), fragment and uses thereof
 JOURNAL Patent: WO 0065067-A 817 02-NOV-2000;
 CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (FR) ; IREMER
 INSTITUT FRANCAIS DE RECHERCHE POUR L'EXPLOITATI; ON DE LA MER (FR)
 FEATURES Location/Qualifiers
 source 1..265118
 /organism="Pyrococcus abyssi"
 /db_xref="taxon:29292"
 /note="Original length of seq 1: 1765118 (for 812 sequences) New sequences: seq 001, from 0.000.001 to 0.349.980, length: 349.980 seq 813, from 0.300.001 to 0.649.980, length: 349.980 seq 814, from 0.600.001 to 0.949.980, length: 349.980 seq 815, from 0.900.001 to 1.249.980, length: 349.980 seq 816, from 1.200.001 to 1.549.980, length: 349.980 seq 817, from 1.500.001 to 1.765.118, length: 265.118"

BASE COUNT 75226 a 61302 c 54005 g 74585 t
 ORIGIN

Query Match 8.1% Score 150.8; DB 6; Length 265118;
 Best Local Similarity 50.8%; Pred. No. 17e-19;
 Matches 527; Conservative 0; Mismatches 487; Indels 24; Gaps 6;

QY 796 GAAGCAAGCTCAGCAAGGAAGAGTCTACTCTCTACCTCCAGATGCGGGGGAAC 855
 DB 59002 GAACCTTATCTTAAAGATGGCGATATACCTTCGCGCAGAGATGGCGCTTAAT 58943
 QY 856 CTCF-----GGCTCTGGACAGGAGCCGAGCGTGTGTATACCGGACCACTGAT 909
 DB 58942 CTGTTCAAGGTTAACTTAGAGGAAGGTTGAAGGTTATTAAGGAGATGAAGCGTT 58883
 QY 910 TACGGGCTTACGATGAGGAGCAAGATGCTCTCATCATGACCGCCAGAGATA 969
 DB 58882 GAGAGCTTTTCCGTTGGATTAATACATTTACAGTCAAGATGCAACATCCGACT 58823
 QY 970 GCGAGCTTACCTCTACGACGCGGAGCTGAACAGTACACCAATACAGGCGGATA 1029
 DB 58822 GAGCTTTACGCTCTCAGGAGATGAAGAGAGATACAGACTTTAAAGCCCTGGA-- 58765
 QY 1030 TTCGGAAGCTCAGACCTTCGACCGAGGACACTTCGCTTCAG--AGCAAGACCTC 1086
 DB 58764 -TTAGGACTTAACTTTCAAAACAGAGCACTTACGCGTACGATGAGGAAG 58706
 QY 1087 GAGATAGAGGCTGCTACCTCAGCGGAGGTTAAAG-----GAGAAGGCCCGGTG 1140
 DB 58705 GAATATAGATGCTGATCATGAGCCGCTTACCTTTAAGCTGGAAGAAATATCTGCA 58646
 QY 1141 ATAGTCTTTCGCCAGCGCGGCGGAGGAGTACGAGACACCGCTTCCTACAGAGATG 1200
 DB 58645 GTGCTGGAATTTACGCTGGGCGCAAAAACAGCTAGGTTATGATTTATACAGAGTTT 58586
 QY 1201 CAGCTGATGGGAGGAGGCTACTACTGCTGCTCTGTAACCCGCGGAGAGAGCGGC 1260
 DB 58585 CAGCTTTTACTTGCACAGAGGCTCTGCTTAAACGCTTTGACTTTATAGATGAAGAGCTT 58409
 QY 1261 TATAGCAGAGACTTCGCGCTCCGCTCAGAGAGACTGCTGAGAGACTTTGAGAGAC 1320
 DB 58525 TATGTCAGAGATTCGG---GACATAGAGGCTCTACACGCTGAGCCGATTTCAAGAC 58469
 QY 1321 ATATAGCAGGATCGAGAGAGTCTTCAAGCTCGAACCGAGCGGAGAGAGCGGCTT 1380
 DB 58468 CTATATGAGGATCTTGAAGAGGCTTTAAACGCTTTGACTTTATAGATGAAGAGCTT 58409
 QY 1381 GGAATTAAGCGGATTAAGCTACGCGGCTTATGACCACTGCGGCTTGAAGAGCGAC 1440
 DB 58408 GCGCTTACCGAGAGATCATGAGGAGGCTTATCAAAATTTGATCGTTGGGCAACAAAC 58349

QY 1441 CTCTTCAGGACGAGATTAAGCAGAGACGCAATAGCTATGCGCTACGAGCTTC 1500
 DB 58348 AGATTAAGAGCTCCCTTAAGCTACGCGCTCAATATCCAGCTGATGCTTCTTGAACG 58289
 QY 1501 TCGGACATAGGCTCTGTGATGACGATGAGGATGCGGCAATCCGTTAGAGAAC--- 1557
 DB 58288 ACCGATATGCGCTATTAATCTCTCCATGATCAGATCGATGCGGATGCTTGAAGTACCA 58229
 QY 1558 GAGAACTCAGAGAGTACGCGCGCTCTCTACGCTCAGACAGTGAAGCGCGGATCTC 1617
 DB 58228 GAGGCTACTGAGAGAGAGAGCCATTAAGTAACTCTCAAAAGTTGAGATGCTCGCTCG 58169
 QY 1618 CTATATCACTCTGCTGAGGACTACCGCTCTCCGCTCAGACAGGCTTATGTTCAACAC 1677
 DB 58168 ATATATCACTCTCAGAGAGGAGACATCCGCTTCTGCTCCGAGGCTTTGCAATTTACATA 58109
 QY 1678 GTGCTCAGAGATGAGGAGAGAGAGGAGGATGAGGATGAGGAGGAGGAGGAGGAGG 1737
 DB 58108 GCGTTGAGATTTTGGGCAAAACCGCTGAGCTGCGATATTCAGATGAGATGATGAC 58049
 QY 1738 CACAGCTCCGCGGAGAGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1797
 DB 58048 TTGAGTATGAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 57989
 QY 1798 TTGAGCGCAAGCTCAG 1815
 DB 57988 TTGAGAGGATGCTGAG 57971

RESULT 13
 BSUB0017/c
 LOCUS
 DEFINITION Bacillus subtilis complete genome (section 17 of 21): from 3197001 to 3414420.
 ACCESSION Z99120 AL009126
 VERSION Z99120.1 GI:2635613
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS

Bacillus subtilis.
 Bacteria: Firmicutes; Bacillales; Bacillaceae; Bacillus.
 1 (bases 1 to 217420)
 Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertoletti, M., Bessieres, P., Bojoltin, A., Borchert, S., Bortis, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C., Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Cummings, N.J., Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D., Emerson, P.T., Enllan, K.D., Erlington, D., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Gallon, N., Ghim, S.Y., Glaser, P., Goffeau, A., Golligorsky, E.J., Grandi, G., Guisepi, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R., Hent, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klaere-Bianchini, M., Klein, C., Kobayashi, Y., Koeller, P., Konigstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, D., Lazarevic, V., Lee, S.M., Levine, A., Liu, H., Masuda, S., Maue, C., Medigue, C., Medina, N., Mellado, R.P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogilwa, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M., Portetle, D., Porro, J.L., Prescott, A.M., Presecan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivola, C., Rocha, E., Roche, B., Rose, M., Sadle, Y., Sato, T., Scanlan, E., Schleich, S., Schroeder, R., Scofield, F., Sekiguchi, J., Sekowska, A., Seror, S.J., Serron, P., Shin, B.S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Trosato, V., Uehyama, S., Vandendol, M., Vannier, F., Vassarotti, A., Viari, A., Wambolt, R., Wedler, E., Wedler, H., Weitzenecker, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yoshida, K., Yoshikawa, H.F., Zumbstein, E., Yoshikawa, H. and Danchin, A.
 The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*

JOURNAL Nature 390 (6657), 249-256 (1997)
MEDLINE 98044033
PUBMED 9384377
2 (bases 1 to 217420)
AUTHORS Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
TITLE Direct Submission
JOURNAL Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48

FEATURES
source

Location/Qualifiers
1. 217420
/organism="Bacillus subtilis"
/strain="168"
/db_xref="taxon:1423"
complement(3. .1277)
/gene="yule"
complement(3. .1277)
/gene="yule"
/function="unknown"
/note="similar to L-rhamnose isomerase"
/codon_start=1
/transl_table=11
/protein_id="CAB5107.1"
/db_xref="GI:2635614"
/db_xref="SPTRMBL:005264"
/translation="MTIKANIDSKAQAYEKWGDIVBEALRQLEQVPISHCWGDDIE
GEVNVNGELSGGIDVGNYPKAPQEPTEBRLDEKALSLPGKRVNLHAYETNRE
AVERDELKPOHEFNENWYMAKNIIGLDENFPLESHEKAADGLTSHDPPIDREFWIR
CIACIRIGEYFEGELGTPCLINIMIPDGKIDIPSDILTPKRLKESLDRJFSEISIO
HNDLSIESKLFEGSSSYVSHPEFYALVTNHKICLIDTGHHPHPEVYSSISNL
LYTDKALAHYSRPRVMSDHYVLDDELRIALAEIYRNHLEKVAIGDPEFDSINRY
MAITIGTRNKRLVLAFLPNSLTLOQEGERTERLALMEEFKTYPRGALWDSICE
OMQVPEKAVLYIKEDQVLLKRASSPIV"
complement(1303. .1617)
/gene="yuld"
complement(1303. .1617)
/gene="yuld"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB5108.1"
/db_xref="GI:2635615"
/db_xref="SPTRMBL:005263"
/translation="MKRKASIMFVHODKYEYQRHDDIPEPAEALKAHGHYSIE
LDEETGRLEFAYLEIEDEEKRMKADTEVCORWKSMAPIKNTSDSPVAIDKEVY
LD"
complement(1635. .3092)
/gene="yulc"
complement(1635. .3092)
/gene="yulc"
/function="unknown"
/note="similar to rhamnulokinase"
/codon_start=1
/transl_table=11
/protein_id="CAB5109.1"
/db_xref="GI:2635616"
/db_xref="SPTRMBL:005262"
/translation="MTYTAIDVAGSSGRINWGLNBEKGLDIOEIHREANGFSQRDHC
LMDIDLKQILOGKVKTLGECVIGIDVAVYLLDEKGDRLRELAISDRRT
DHTIDLKLEHTSKAAYOKTIOPOPNTYOLFEEDRELLKTKDKIMIPDLYCYL
TGRAVETITVSTTOLINVTGNLDPLEALVATLQOPALPEPCGECGLKLNEMPP
DYDLPACKWTVATHTDASAVIAPGVNCGWATISSGWSLIGVENKTPITITLALEN
NNTNERNANTTIRPLANKIIGMWVIOEQKQOLADYSFOOLAEBAKTEPPOFTINLD
KRLPENMNIKEIOHCROTROKIPRYAGELACCIYSNLAIYIAIKELLETTERKI
EOPHIIGGARGNDFLNOLITADMSKAVYAGPIEATATGNLMOMIAAEVKDKIEAKO
VVRNSPEIKFTPEKIDRSRTIOSFOQTVKALSK"
complement(3097. .3873)
/gene="yulb"
complement(3097. .3873)
/gene="yulb"
/gene="yulb"

gene
CDS

/function="unknown"
/note="similar to transcriptional regulator (Deor family)"
/codon_start=1
/transl_table=11
/protein_id="CAB5110.1"
/db_xref="GI:2635617"
/db_xref="SWISS-PROT:005261"
/translation="MTVAERQKQIYIVNMRSSIRYSELDFSYTEETIRDDLEKE
KEKRLSRHOGAVSIOQSEIHFSERETNTYERKAIAHEBAKYKSGDRIILDAST
TANYMAKILPDIELTVITNSMAKALISKNENISVISTGILLLESLSAGPLAERSL
EYHVNRKTELSCKGFDVHNGMSDNEMOALLKRMERSDQTLNADSSKMGNRRSH
IASLDVSRRLITDSDGLDPASVAKLEDKKVTAVPLSKRG"
complement(3930. .5999)
/gene="yuxg"
complement(3930. .5999)
/gene="yuxg"
/function="unknown"
/note="alternate gene name: yula
similar to sorbitol-6-phosphate 2-dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="CAB5111.1"
/db_xref="GI:2635618"
/db_xref="SWISS-PROT:P40747"
/translation="WVKHIMDSERAAQKRGVELYRSNLIGSDRTVCNMGGNTSM
KTEKDPGRGRIETVMYKSGSDLAHMAKNAHNSGLTLDIPIRLIKRQDPDEMUYL
SHCMIDSKHPRIETLHAFLPYKHVDTHHPALITSCADNGKOIADITYGNRPVW
VPYVRPGLTISKMAIAGVANNPAAELVLEKRLVTWGETSEYCTQITSIIOEAYQ
INRINOHVEFGSKAROPLEDRKROILAGIMVINGAASEKKMLISYDDHDVLEF
VNSVQALDSQIOIACPDHLYHTKRPVLLDMNPEPOVHKLADLKSVEVEFTSEYQ
AYTRNOQDLPDESAPRYILLTGIGVNTGKSYAMSVSGALYRRALVAMKGATLQ
GQVYIAHENSRYVEVETLEYLTLLAPPEAPERSRVALITGGAGIGSAAORRFAE
GGHYIYADNLIEBQKIAEINDAYKGRAMVKNVYTEEDYQSFERRAALAYGID
IVNNGSLATSPDEFSLKNNLNKNNVJGTFVLARAFKQMKHONKGSVAFVS
KNSVYAGKNAASVSKALETHLARCILABGGEGFIRVNSVLPDAVLQGSALWSSWR
EERRAAVGIPEPDLEBHYRKRKTLVNIYIPEDIAEIAEFASSAKAKETGCMITVDG
VPAAFTP"
complement(6109. .8125)
/gene="tlpb"
complement(6109. .8124)
/gene="tlpb"
complement(6137. .8125)
/gene="tlpb"
/codon_start=1
/transl_table=11
/product="methyl-accepting chemotaxis protein"
/protein_id="CAB5112.1"
/db_xref="GI:2635619"
/db_xref="SWISS-PROT:P39217"
/translation="MGKFIQWIKOPSISKPLIAFLAVILPVGVLAFYSOSANAL
DRELISAKNGVFNLSSTONKLEDKKADIDYSEVVDNDIILGKKKTLTKERFOYT
TUNDGAIITAASEDKKLYKYPDSGYPKGFDPGRGRWYQNAVEKQANFSEPYDEA
TGDIVATISKQLDKSGVIALDLNIDELVLAASRIKIGKEGFAFLTYGKKKTIHPTI
KPGTSGDWTNQVYSKRKEGSEFEYERKEKKAFTNKLATGKLAGTIVSELADAS
SPVLANVAIILICVIVIGILYIIIRATKYLPKRLVSTSAKISGDLVEVILDSRN
EPGOLGSEFNEMASLSRVYGIOTSEVNAVSSEELTSAQTSKATHTILAEQF
SDGNEAOSEKLETSSNHTISOMNGISKVAAQASTTIRKSSIOSSEAGSEKELYEVG
QMKTIQOSYOKAAVYKGLTYSODITSLYINGIADONTLAINALAEARAGEYQ
KGRSVAAEYKRLAVOSADSAKEIBGLIDETVAEISTLSMFOVNHREKREGIQTIDQ
TASFSFOIYEMTQIISGELONLNAVTEQLSAGQEVSSAVEDISAVAKKSSAGTODIA
ASAEBOIASMEELSSSAETLANNAEELODITKFKKIES"
complement(8204. .10224)
/gene="mcpa"
complement(8204. .8228)
/gene="mcpa"
complement(8239. .10224)
/gene="mcpa"
/function="chemotaxis deficiency toward glucose and
alpha-methyl-glucoside"
/codon_start=1
/transl_table=11
/product="methyl-accepting chemotaxis protein"
/protein_id="CAB5113.1"

/db_xref="GI:2635620"
 /db_xref="SWISS-PROT:P39214"
 /translation="MKRILOIKORSTRLKLFSLILIPVILAEVQSSSL
 DROMGSALENOVQDLINISIGKENSADYEEMLTKEKYNASSTREKSOY
 SINKVESITSDKGFHETRPDLPMSEGNVPVRMDIKRVANKGVITDPKITS
 TMTPTVIAAOTKDSGVIAINMTIENMLTKTKVNGTGTATIMKRDVVAHPNE
 OSGETELMDLMDKLSADKDPQYTMGDGDKRKAFDNKLGMTIGMTIDEIHEA
 QPVHLALIVLAAIIGIIVMTLIRSIITTPKQLVSSKRISEGLTETIDIRSD
 ELBELGSPNNMASLSRLIHAIDSVDNVAASEELTASAOTSKAPETHITLIDF
 SNGERKENNETIETAEHIIYOMDGLTMAQASEVITDSVOSTEIASEGGLVHOTG
 OMNYIDKSEAEQVRLGFTSKSDINILIRVINGINDQRIHLALNAIEARAGEY
 RPSVVAERKLAQVSDSAKEIEGLITIEVKEINISLGFOSVNEVGTIDITDK

Query Match 8.0%; Score 150.2; DB 1; Length 217420;
 Best Local Similarity 52.3%; Pred. No. 2.2e-19;
 Matches 357; Conservative 0; Mismatches 323; Indels 3; Gaps 1;

1135 CCGGTGATGCTTCGTCACAGGGGGCCGAGGCGCATGTACAGACCCGCTGCTAC 1194
 113504 CCACTTATCTTAAACATACACGGGGGTCCTATATGATGATACGACATATTTTCAT 113445
 1195 GAGATGAGCTGATGAGGCAAGGCGCTACTACTGCTGCTGGAACCGCGGGGAGC 1254
 113444 GAGTTTCAGGTGCTGGGGGGAAGATAGCGGCTCTTATATCAATCCGAGAGAGC 113385
 1255 GAGGCTATAGCGAAGCTTCGCGCTCCGCTCTGAGAGAGAGCTTGAGAGACTTT 1314
 113384 CACGGCTACGGGAGGATTTTGAATGCGGTACAGAGATTTATGAGGGAAGATTAAT 113325
 1315 GAGGACATATGAAAGCGCATGAGAGATTTCTCAAGCTGACACCGGCGGAGAGAG 1374
 113324 GAGGATGTGATGACAGGCTGTGATGAGGCTATCAACGAGATCCGCAATTTGATCTTAAG 113265
 1375 CGGCTTGGAATTAACGGGCATTAAGCTACGGGCTTCAACCAACTGGGCTTGACTAC 1434
 113264 CGGCTGCTGTCACGGGCGGAAGCTACGGAGGTTTATGACCAACTGATGCTGCGGAG 113205
 1435 AGGACCTCTTCAAGGAGAAATTAAGCGAAGAGGCTTAAGTACTGCTCACCAGCTAC 1494
 113204 ACAGACCGCTTAAAGCTGCGGTACCCAGCGCTCGATCAATTAATGATCAGCTTTAC 113145
 1495 GCGTCTCGACATAGGGCTCTGTGACAGCTGAGGTCATCGGGCAATTCGTTAGAG 1554
 113144 GGGGTGAGATGATGCGGCTATTTCTTACAGACTGGCAGCTTGAACATGTTTGAAG 113085
 1555 AAC---GAGAACTTCAGGAAGCTCAGCCGCTGTTTACGCTCAGAACGTGAAGCGCCG 1611
 113084 GACACAGAAAAGCTCTGGGACCGGCTCTCTTAAATATAGCAGCAAAAGTGAAGACCG 113025
 1612 ATACTCCTTAATCCACTGCTTGAGACTACCGCTGTCCGCTGACACAGAGCTTATGTTTC 1671
 113024 CTTTATATCTGATGCGGAGCGGAGTGAACGATGCGCGATCGAGAGGCGGAGAGCTG 112965
 1672 TACACGTGCTCAAGACATGGGCAAGAAAGCTTACATAGCAGATTTCAAGCGCGGCGCC 1731
 112964 TTTATGCTCTGAAAAAATGGGCAAGAAACCAACCTTGTCGTTTCCGATGATGATG 112905
 1732 CAGGCGCACAGCGTCCGCGGAAGCCCGAGGACACAGCGGCTTACAGGCTCTTCAT 1791
 112904 CACAATTTATACAGCGCACCGGACACCAAGACAGCGGATCAAGGCTGATTTATATCAGC 112845
 1792 GAGTCTTCGAGGCGAAGCTCA 1814
 112844 TCATGCTTGTATCAACATCTCTA 112822

RESULT 14
 AX006271 1971 bp DNA linear PAT 24-AUG-2000
 LOCUS
 DEFINITION Sequence 1 from Patent W09903984.
 ACCESSION AX006271
 VERSION AX006271.1 GI:9929124
 KEYWORDS

SOURCE
 ORGANISM Bacillus subtilis.
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 REFERENCE 1 (bases 1 to 1971)
 AUTHORS Estell,D.A.
 TITLE Proteases from gram-positive organisms
 JOURNAL Patent: WO 9903984-A 1 28-JAN-1999;
 GENENOR INTERNATIONAL B.V. (NL); ESTELL DAVID A (US)
 FEATURES
 SOURCE location/Qualifiers
 1..1971
 /organism="Bacillus subtilis"
 /db_xref="taxon:1423"

BASE COUNT 558 a 431 c 524 g 458 t
 ORIGIN

Query Match 8.0%; Score 149.8; DB 6; Length 1971;
 Best Local Similarity 52.3%; Pred. No. 2.2e-19;
 Matches 356; Conservative 0; Mismatches 322; Indels 3; Gaps 1;

1135 CCGGTGATGCTTCGTCACAGGGGGCCGAGGCGCATGTACAGACCCGCTGCTAC 1194
 1291 CCACTTATCTTAAACATACACGGGGTCCCATATGATGATACGACATATTTTCAT 1350
 1195 GAGATGAGCTGATGAGGCAAGGCGCTACTACTGCTGCTGGAACCGCGGCGAGC 1254
 1351 GAGTTTCAGGTGCTGGGGGGAAGATAGCGGCTCTTATATCAATCCGAGAGAGAGC 1410
 1255 GAGGCTATAGCGAAGCTTCGCGCTCCGCTCTGAGAGAGAGCTGCTTGAGAGACTTT 1314
 1411 CACGGCTACGGGAGGATTTTGAATGCGGTACAGAGATTTATGAGGGAAGATTAAT 1470
 1315 GAGGACATATGAAAGCGCATGAGAGATTTCTCAAGCTCAGAACCGGCGGAGAGAG 1374
 1471 GAGATGTGATGACAGGCTGTGATGAGGCTATCAACGAGATCCGATATTTGATCTTAAG 1530
 1375 CGGCTTGGAATTAACGGGCATTAAGCTACGGGCTTCAACCAACTGGGCTTGACTAC 1434
 1531 CGGCTGCTGTCACGGGCGGAAGCTACGGAGGTTTATGACCAACTGATGCTGCGGAG 1590
 1435 AGGACCTCTTCAAGGAGGATTAAGGAGAAAGGCAATTAAGTACTGCTCACCAGCTAC 1494
 1591 ACAGACCGCTTAAAGCTGCGGTACCCAGCGCTCGATCAATTAATGATCAGCTTTAC 1650
 1495 GCGTCTCGACATAGGGCTCTGTGACAGCTGAGGTCATCGGGCAATTCGTTAGAG 1554
 1651 GGGGTGAGATGATGCGGCTATTTCTTACAGACTGGCAGCTTGAACATGTTTGAAG 1710
 1555 AAC---GAGAACTTCAGGAAGCTCAGCCGCTGTTTACGCTCAGAACGTGAAGCGCCG 1611
 1711 GACACAGAAAAGCTCTGGGACCGGCTCTCTTAAATATAGCAGCAAAAGTGAAGACCG 1770
 1612 ATACTCCTTAATCCACTGCTTGAGACTACCGCTGTCCGCTGACACAGAGCTTATGTTTC 1671
 1771 CTTTATATCTGATGCGGAGCGGAGTGAACGATGCGCGATCGAGAGGCGGAGAGCTG 1830
 1672 TACACGTGCTCAAGACATGGGCAAGAAAGCTTACATAGCAGATTTCAAGCGCGGCGCC 1731
 1831 TTTATGCTCTGAAAAAATGGGCAAGAAACCAACCTTGTCGTTTCCGATGATGATG 1890
 1732 CAGGCGCACAGCGTCCGCGGAAGCCCGAGGACACAGCGGCTTACAGGCTCTTCAT 1791
 1891 CACAATTTATACAGCGCACCGGACACCAAGACAGCGGATCAAGGCTGATTTATATCAGC 1950
 1792 GAGTCTTCGAGGCGAAGCTC 1812
 1951 TCATGCTTGTATCAACATCTCTC 1971

RESULT 15
 AX431809 1958 bp DNA linear PAT 28-JUN-2002
 LOCUS
 DEFINITION Sequence 224 from Patent W00229113.
 ACCESSION AX431809

